1965

The utilization of concomitant information in sequential procedures for the comparison of two treatments

Thomas Dean Roseberry
Iowa State University

ROSEBERRY, Thomas Dean, 1937–
THE UTILIZATION OF CONCOMITANT INFORMATION IN SEQUENTIAL PROCEDURES FOR THE COMPARISON OF TWO TREATMENTS.

Iowa State University of Science and Technology
Ph.D., 1965
Mathematics

THE UTILIZATION OF CONCOMITANT INFORMATION IN SEQUENTIAL PROCEDURES FOR THE COMPARISON OF TWO TREATMENTS

by

Thomas Dean Roseberry

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1965

TABLE OF CONTENTS

I. INTRODUCTION

A. Sequential Medical Trials

Armitage (3) writes,

> A striking aspect of post-war research in clinical and preventive medicine is the extent to which new drugs and vaccines have been tested by controlled trials. The essential feature of these trials has been the random allocation of subjects to groups receiving different treatments, a principle of experimentation due to Sir Ronald Fisher and now regarded as fundamental in many fields of scientific research....The statistical methods for testing the significance of observed differences between groups, normally applied at the end of such a trial, assume that the number of subjects included was determined either before the start of the trial, or at least by some circumstance independent of the results themselves....If a trial is brought to a close at a certain point because, at that point, the results for the different treatments are either sufficiently alike or sufficiently distinct, then the traditional tests of significance are, strictly speaking, inapplicable. In prophylactic and therapeutic trials, however, there are strong ethical reasons why the results should, whenever possible, be subject to continuous analysis as they become available; it is clearly desirable that a trial should be discontinued as soon as it can safely be concluded that one treatment is more effective than another. Furthermore, quite apart from ethical considerations, this 'sequential' method of experimentation can be defended on grounds of economy, since it leads to a reduction in the average number of subjects required.

Bross (8) suggests another reason why sequential techniques may be particularly appropriate in clinical medical experimentation. A disease to be studied may be relatively rare so that in order to acquire a reasonable series of patients it may be necessary to continue the study over months or even years. Sequential techniques are particularly pertinent in such situations where observations are gathered overtime with plenty of time for examination of results.

Let us now turn to the general problem of design in sequential comparative trials. Armitage (2, pp. 4,5) makes a broad distinction between two types of designs — those in which treatment comparisons are made between subjects, because each subject is given only one of the alternative treatments, and those in which comparisons are made within subjects due to the fact that each subject is given more than one treatment. He further points out that in the former kind of design (involving between-subject comparisons), some increase in precision may result by grouping together subjects whose responses might be similar — say for reasons of age, sex, or severity of the disease — a technique which we shall call "stratification." Comparisons are then made between subjects within a particular grouping or stratum.

This idea may be carried further. Suppose a useful measured concomitant observation can be made prior to administering a treatment. By a useful measured concomitant observation we mean a quantitative observation which is reasonably predictive of the subject's behavior in absence of treatment. For example, it may be an observation of some characteristic such as pre-treatment blood pressure in a study to ascertain a treatment's ability to alleviate hypertension, so long as these observations and the responses to treatment are highly correlated. Although stratification could be imposed on the concomitant observations, one might expect increased precision by direct sequential utilization of these observations as in the fixed sample analogue, the analysis of covariance. This is suggested because with stratification one must make the implicit assumption that all concomitant observations within a stratum are identical. This assumption could be made more tenable by further refinement of the stratification; however this would pose an additional problem in sequential analysis. Most available methods are conveniently performed on pairs of observations, each member of each pair being randomly allocated to one of the two treatments. Hence excessive refinement of the stratification may result in wastage since at any particular stage of investigation an appreciable number of patients may be unpaired because they have all fallen into different strata. This consequence can be of great importance if one is studying a relatively rare disease so that patients are entering a study over a long period of time, especially if there is a gradual trend in response throughout a trial, affecting both treatment groups equally. Armitage (2, p. 6) suggests that such a trend could occur for environmental reasons or because either the standard of assessment of response or the nature of the disease in different patients is gradually changing. The difference between two observations due to the trend effect is of course minimized when the observations are successive ones. Therefore, the pairing of successive entrants, as would obtain with direct utilization of concomitant information, should diminish the trend effect whatever its cause. But with stratification, where the basis for stratification is something other than time of arrival, pairing must be within strata so that the members of any particular pair are unlikely to be successive entrants. Indeed they may enter the study at widely differing points in time.

The problem of stratification requires further discussion. In the framework from which the preceding remarks arose, direct utilization of concomitant information when a useful covariate is available was proposed as an alternative to using this covariate as a basis for stratification. Pairing of observations is common to both approaches, but in the covariance approach successive entrants into the trial are paired; with stratification, successive entrants into a stratum are paired. But it is noted that direct utilization of concomitant information is possible in conjunction with stratification. A qualitative factor which is a basis for stratification may be present in addition to a useful covariate. In this case, the covariance technique would be used within the strata, i.e. successive entrants into a stratum would be paired. But if stratification is imposed on the covariate as well as on the qualitative factor, two dimensional stratification is involved and the pairing is of successive entrants into the intersection of a stratum with the covariate as its basis and a stratum with the qualitative factor as its basis.

Another point should be raised in passing. One possible basis for stratification might be time of entry, i.e. the first say 2k entrants might be taken as the first stratum, the next 2k patients as the second stratum, etc. Then successive entrants into a stratum, which are also successive entrants into the trial in this situation, could be paired (or the pairing could be done at random within each stratum) with the underlying rationale that "neighbors" are likely to be similar. And it is possible that a useful covariate is present which is in some way related to time of entry so that the responses of "neighbors" are likely to be correlated. In the statement of the problem given in the following section we shall not consider this situation but rather that in which the covariate is not related to time of entry and in which patients enter the trial in a random manner so that the responses of successive entrants are uncorrelated.

B. Statement of the Problem

In this thesis we shall consider the problem of utilizing measured concomitant information in sequential procedures for the comparison of two treatments. The treatments may be two drugs, two diets, two operative techniques, etc., perhaps with treatment 1 being a "new" treatment and treatment 2 being a "standard" or "control" treatment. It is assumed that subjects enter the trial at random and that successive subjects are paired. One member of each pair is allocated at random to treatment 1 or 2, the other member being assigned to the remaining treatment. We shall designate our measured concomitant variate as x and the response to treatment, which is also taken to be quantitative, shall be designated as y. Then y_{ij} denotes the response of the subject in the j-th pair receiving the i-th treatment. Similarly, x_{ij} denotes the concomitant observation of the subject in the j-th pair to receive the i-th treatment.

In the j-th pair we form the treatment 1 minus treatment 2 response difference,

$$y_j = y_{1j} - y_{2j} \quad ,$$

and the corresponding concomitant observation difference,

$$x_j = x_{1j} - x_{2j} \quad .$$

It is assumed that y_j, x_j follow the bivariate normal distribution law, with means μ_x, μ_y, variances σ_x^2, σ_y^2, and correlation coefficient ρ.

The joint probability density function (p.d.f.) of x, y is

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

$$\exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right] \quad (1.1)$$

Alternatively we may express f(x,y) as the product the marginal p.d.f. of x and the conditional p.d.f. of y (given x). Then 1.1 becomes

$$f(x,y) = g(x)\ h(y|x) = N(x;\mu_x,\sigma_x^2)\ N(y;\mu_{y|x},\sigma_{y|x}^2)$$

$$= \left[\frac{1}{\sigma_x(2\pi)^{1/2}} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right]\right] \left[\frac{1}{\sigma_{y|x}(2\pi)^{1/2}} \exp\left[-\frac{(y-\mu_{y|x})^2}{2\sigma_{y|x}^2}\right]\right]$$

$$= \frac{1}{2\pi\sigma_x\sigma_{y|x}} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_{y|x})^2}{2\sigma_{y|x}^2}\right], \quad (1.2)$$

where

$$\sigma_{y|x}^2 = \sigma_y^2\ (1-\rho^2)\ , \quad (1.3)$$

$$\mu_{y|x} = \mu_y + \beta(x-\mu_x)\ , \quad (1.4)$$

and

$$\beta = \rho\,\frac{\sigma_y}{\sigma_x} \quad (1.5)$$

Since the allocation of subjects to treatments is random, we have

$$\mu_x = E(x_j) = E(x_{1j} - x_{2j}) = 0 \ . \tag{1.6}$$

Hence, our basic model assumptions are that

$$x_j \sim NI(x;\ 0,\sigma_x^2) \ , \tag{1.7}$$

and

$$\begin{aligned} y_j|x_j &\sim NI(y;\ \mu_y + \beta x_j, \sigma_{y|x}^2) \\ &= NI(y;\ \mu + \beta x_j, \sigma^2) \ , \end{aligned} \tag{1.8}$$

where

$$\mu = \mu_y = E(y_j) = E(y_{1j} - y_{2j}) = \mu_1 - \mu_2 \tag{1.9}$$

$$\beta = \frac{\rho\sigma_y}{\sigma_x} \ , \tag{1.10}$$

and

$$\sigma^2 = \sigma_y^2(1-\rho^2) = \sigma_{y|x}^2 \tag{1.11}$$

For future reference we note that 1.8 may be equivalently written as

$$y_j|x_j = \mu + \beta x_j + e_j \ , \tag{1.12}$$

where

$$e_j \sim NI(e;\ 0,\sigma^2) \tag{1.13}$$

We shall attempt to develop sequential tests for discrimination between two treatments in each of two distinct hypothesis formulations. In what will be termed as formulation 1, we shall test the hypothesis $H_T: \mu = \mu_T$ against the hypothesis $H_A: \mu = \mu_A$, where we are using "T" and "A" to denote "test" and "alternative," respectively. In formulation 2, we shall test the hypothesis $H_T: \mu = \mu_T$ against the hypothesis $H_A: \mu = \mu_T + \gamma\sigma$ (γ is a specified constant). Formulation 1 is appropriate if certain absolute changes in the mean are of interest, whatever the value of σ. Formulation 2 is appropriate when one wishes to detect when the mean differs from μ_T by at least γ standard deviation units.

II. A LARGE SAMPLE TEST UTILIZING CONCOMITANT INFORMATION

A. The Sequential Probability Ratio Test (SPRT)

In this section we shall present only a brief description of the sequential probability ratio test (s.p.r.t.) since a thorough discussion of this important procedure is available in Wald (28, pp. 37-48).

Suppose we take a sample of n successive observations $x_1, x_2, \ldots, x_n$ from a probability density function (p.d.f.) $f(x;\theta)$ (The s.p.r.t. also holds for the discrete probability law). Then the ratio of the probabilities of the sample on the hypotheses $H_T: \theta=\theta_T$ and $H_A: \theta=\theta_A$

$$P_n = \frac{\prod_{i=1}^{n} f(x_i;\theta_A)}{\prod_{i=1}^{n} f(x_i;\theta_T)} \tag{2.1}$$

Two positive constants C,D (D<C) are chosen. The s.p.r.t. for testing the simple hypothesis H_T against the simple hypothesis H_A is conducted as follows:

On the n-th stage P_n is computed. If

$$D < P_n < C \tag{2.2}$$

another observation is taken. If

$$P_n \geq C \tag{2.3}$$

the process is terminated by accepting H_A. If

$$P_n \leq D \tag{2.4}$$

the process is terminated by accepting H_T.

The constants C, D are chosen so that the test will have prescribed type I and type II errors, i.e., so that

$$P(\text{reject } H_T \mid H_T) = \alpha_1$$

and (2.5)

$$P(\text{accept } H_T \mid H_A) = \alpha_2$$

Wald (28, p. 41) derives the important inequalities

$$C \leq \frac{1-\alpha_2}{\alpha_1}$$

and (2.6)

$$D \geq \frac{\alpha_2}{1-\alpha_1}$$

Due to the essentially discrete behavior of the test, rarely are the boundaries exactly attained. However, if the boundaries are such that C and D are exactly attained when attained at all, we have

$$C = \frac{1-\alpha_2}{\alpha_1}$$

and (2.7)

$$D = \frac{\alpha_2}{1-\alpha_1}$$

Further, Wald demonstrates that no serious error results from using 2.7 to define C, D if α_1 and α_2 are relatively small, as is usually the case in practice. Hence, for practical purposes we may conduct Wald's s.p.r.t. as defined by Equations 2.1 - 2.4 and 2.7.

An additional point to be made is that it is often much easier computationally to conduct the s.p.r.t. by taking logarithms, i.e. by computing log P_n with boundaries (log C, log D).

The startling feature of this sequential test when compared with fixed sample number tests is that the s.p.r.t. can be conducted without solving any distributional problems as are necessary in conventional testing (in which case the distribution of the test statistic must be found). As Wald (28, p. 48) has pointed out, "Distribution problems arise in connection with the sequential process only if it is desired to find the probability distribution of the number of trials necessary for reaching a final decision. But this is of secondary importance as long as we know that the sequential test on the average leads to a saving in the number of observations."

B. A Large Sample SPRT, Formulation 1

Wald's s.p.r.t. gives a convenient and practical procedure for discriminating between simple hypotheses. In the formulations of the problem which we are here considering, however (see Section B, Chapter I), the test and alternative hypotheses, H_T and H_A, are composite since β, σ_x^2, and σ^2 are unspecified. We consider now a large sample theory for dealing with composite hypotheses which has been suggested by Cox (10). In this section we shall be concerned with formulation 1 of the problem, i.e.

$$H_T:\mu = \mu_T \quad \text{versus} \quad H_A:\mu = \mu_A \ .$$

Assume for the moment that β, σ_x, and σ^2 are known to be β_o, σ_{xo}^2, σ_o^2 . Then a solution to our problem is given by Wald's s.p.r.t. Thus, after taking n observations (x_i, y_i) from $f(x_i, y_i)$, our test would be based upon

$$\log P_n = \frac{\log \prod_{i=1}^{n} f(x_i,y_i;\mu_A,\sigma_{xo}^2,\beta_o,\sigma_o^2)}{\log \prod_{i=1}^{n} f(x_i,y_i;\mu_T,\sigma_{xo}^2,\beta_o,\sigma_o^2)} \tag{2.8}$$

$$= L_n(\mu_A,\theta_o) - L_n(\mu_T,\theta_o) \ ,$$

where θ denotes $(\sigma_x^2,\beta,\sigma^2)$, and

$$L_n(\mu,\theta) = \log \prod_{i=1}^{n} f(x_i,y_i;\mu,\theta). \tag{2.9}$$

Here it is seen from Equations 1.1 — 1.11 that

$$f(x,y;\mu,\theta) = \frac{1}{2\pi\sigma_x\sigma} \exp\left[-\frac{x^2}{2\sigma_x^2} - \frac{(y-\mu-\beta x)^2}{2\sigma^2}\right]. \tag{2.10}$$

The approach of Cox (10) is to consider a large sample theory in which the log likelihoods are expanded as far as quadratic terms and in which $\mu_T - \mu$ and $\mu_A - \mu$ are taken to be of order $n^{-1/2}$. We shall retain this latter feature, but we shall consider exact Taylor expansions of the log likelihoods. A rationale for treating $\mu_T - \mu$ and $\mu_A - \mu$ as being of order $n^{-1/2}$ is discussed at the end of this chapter.

We have for $j = T,A$,

$$L_n(\mu_j,\theta_o) = L_n(\mu,\theta_o) + (\mu_j-\mu)\left[\frac{\partial L_n(\mu_j,\theta_o)}{\partial \mu_j}\right]_{\mu_j=\mu} + \frac{(\mu_j-\mu)^2}{2}\left[\frac{\partial^2 L_n(\mu_j,\theta_o)}{\partial \mu_j^2}\right]_{\mu_j=\mu} + \frac{(\mu_j-\mu)^3}{6}\left[\frac{\partial^3 L_n(\mu_j,\theta_o)}{\partial \mu_j^3}\right]_{\mu_j=\bar{\mu}}, \qquad (2.11)$$

where $\bar{\mu}$ is some point lying between μ_j and μ. From 2.10 we note that

$$\frac{\partial^3 L_n(\mu_j,\theta_o)}{\partial \mu_j^3} = 0 \quad . \qquad (2.12)$$

Therefore the expansion of 2.8 about μ is

$$L_n(\mu_A,\theta_o)-L_n(\mu_T,\theta_o) = (\mu_A-\mu_T)\frac{\partial L_n(\mu,\theta_o)}{\partial \mu} + \tfrac{1}{2}(\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)\frac{\partial^2 L_n(\mu,\theta_o)}{\partial \mu^2}, \qquad (2.13)$$

where we are letting

$$\left[\frac{\partial L_n(\mu_T,\theta_o)}{\partial \mu_T}\right]_{\mu_T=\mu} = \left[\frac{\partial L_n(\mu_A,\theta_o)}{\partial \mu_A}\right]_{\mu_A=\mu} , = \frac{\partial L_n(\mu,\theta_o)}{\partial \mu} \quad (2.14)$$

and $\frac{\partial^2 L_n(\mu,\theta_o)}{\partial \mu^2}$ have the analogous meaning.

In practice θ will not be known, and we shall assume that no prior probability distribution of θ is available. It is plausible to consider instead of 2.8

$$L_n(\mu_A,\hat{\theta}) - L_n(\mu_T,\hat{\theta}) , \quad (2.15)$$

where $\hat{\theta}$ denotes the maximum likelihood (m.l.) estimator for θ for samples of size n (actually we shall use m.l. estimators which have been corrected for bias where necessary).

Expanding 2.15 about μ we obtain, from 2.13,

$$(\mu_A-\mu_T)\frac{\partial L_n(\mu,\hat{\theta})}{\partial \mu} + \tfrac{1}{2}(\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)\frac{\partial^2 L_n(\mu,\hat{\theta})}{\partial \mu^2}. \quad (2.16)$$

We now expand 2.16 about σ_x^2. From 2.10 we see that for $s \geq 1$ and any $r \geq 1$,

$$\frac{\partial^{r+s} L_n(\mu,\sigma_x^2,\hat{\beta},\hat{\sigma}^2)}{\partial \mu^r \, \partial(\sigma_x^2)^s} = 0 \quad (2.17)$$

Thus, the expansion of 2.16 about σ_x^2 is given by

$$(\mu_A-\mu_T)\frac{\partial L_n(\mu,\sigma_x^2,\hat{\beta},\hat{\sigma}^2)}{\partial\mu}$$

$$+ \tfrac{1}{2}(\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)\frac{\partial^2 L_n(\mu,\sigma_x^2,\hat{\beta},\hat{\sigma}^2)}{\partial\mu^2} \qquad (2.18)$$

We now expand 2.18 about β. Here we see that for $r > 1$,

$$\frac{\partial^{r+1} L_n(\mu,\sigma_x^2,\beta,\hat{\sigma}^2)}{\partial\mu\;\partial\beta^r} = 0, \qquad (2.19)$$

and for $r \geq 1$

$$\frac{\partial^{r+2} L_n(\mu,\sigma_x^2,\beta,\hat{\sigma}^2)}{\partial\mu^2\;\partial\beta^r} = 0\ , \qquad (2.20)$$

Therefore, expanding 2.18 about β, we obtain

$$(\mu_A-\mu_T)\frac{\partial L_n(\mu,\sigma_x^2,\beta,\hat{\sigma}^2)}{\partial\mu} + (\mu_A-\mu_T)(\hat{\beta}-\beta)\frac{\partial^2 L_n(\mu,\sigma_x^2,\beta,\hat{\sigma}^2)}{\partial\mu\;\partial\beta}$$

$$+ \tfrac{1}{2}(\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)\frac{\partial^2 L_n(\mu,\sigma_x^2,\beta,\hat{\sigma}^2)}{\partial\mu^2} \qquad (2.21)$$

Finally, our expansion of 2.15 may be obtained by expanding 2.21 about σ^2. Hence, an exact Taylor expansion of 2.15 about (μ,θ) is given by

$$
\begin{aligned}
L_n(\mu_A,\hat{\theta})-L_n(\mu_T,\hat{\theta}) = {} & (\mu_A-\mu_T)\frac{\partial L_n(\mu,\theta)}{\partial\mu} + (\mu_A-\mu_T)(\hat{\beta}-\beta)\frac{\partial^2 L_n(\mu,\theta)}{\partial\mu\,\partial\beta} \\
& + \tfrac{1}{2}(\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)\frac{\partial^2 L_n(\mu,\theta)}{\partial\mu^2} \\
& + (\mu_A-\mu_T)(\hat{\sigma}^2-\sigma^2)\frac{\partial^2 L_n(\mu,\theta)}{\partial\mu\,\partial(\sigma^2)} \\
& + \tfrac{1}{2}(\mu_A-\mu_T)(\hat{\sigma}^2-\sigma^2)^2\,\frac{\partial^3 L_n(\mu,\sigma_x^2,\beta,\sigma_1^2)}{\partial\mu\;\partial(\sigma_1^2)^2} \\
& + (\mu_A-\mu_T)(\hat{\beta}-\beta)(\hat{\sigma}^2-\sigma^2)\,\frac{\partial^3 L_n(\mu,\theta)}{\mu\,\partial\beta\,\partial(\sigma^2)} \\
& + \tfrac{1}{2}(\mu_A-\mu_T)(\hat{\beta}-\beta)(\hat{\sigma}^2-\sigma^2)^2\,\frac{\partial^4 L_n(\mu,\sigma_x^2,\beta,\sigma_2^2)}{\partial\mu\,\partial\beta\;\partial(\sigma_2^2)^2} \\
& + \tfrac{1}{2}(\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)(\hat{\sigma}^2-\sigma^2)\,\frac{\partial^3 L_n(\mu,\theta)}{\partial\mu^2\;\partial(\sigma^2)} \\
& + \tfrac{1}{4}(\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)(\hat{\sigma}^2-\sigma^2)^2\,\frac{\partial^4 L_n(\mu,\sigma_x^2,\beta,\sigma^2)}{\partial\mu^2\;\partial(\sigma_3^2)^2}\,, \qquad (2.22)
\end{aligned}
$$

where $\sigma_i^2(i=1,2,3)$ denote some points lying between $\hat{\sigma}^2$ and σ^2.

Now the test when θ is known is based on 2.8. A test in which θ is replaced by $\hat{\theta}$ is based on 2.15. Following Cox (10) we compare the Taylor expansions of 2.8 and 2.15, which are given in 2.13 and 2.22 respectively, and see that the Wald-type test wherein θ is replaced by $\hat{\theta}$ is asymptotically equivalent to the s.p.r.t (with θ known) if and only if

$$\sum_{i=1}^{7} X_i \xrightarrow{(p)} 0 , \tag{2.23}$$

where

$$X_1 = (\mu_A-\mu_T)(\hat{\beta}-\beta) \frac{\partial^2 L_n(\mu,\theta)}{\partial\mu\, \partial\beta} , \tag{2.24}$$

$$X_2 = (\mu_A-\mu_T)(\hat{\sigma}^2-\sigma^2) \frac{\partial^2 L_n(\mu,\theta)}{\partial\mu\, \partial(\sigma^2)} , \tag{2.25}$$

$$X_3 = (\mu_A-\mu_T)(\hat{\sigma}^2-\sigma^2)^2 \frac{\partial^3 L_n(\mu,\sigma_x^2,\beta,\sigma_1^2)}{\partial\mu\, \partial(\sigma_1^2)^2} , \tag{2.26}$$

$$X_4 = (\mu_A-\mu_T)(\hat{\beta}-\beta)(\hat{\sigma}^2-\sigma^2)\frac{\partial^3 L_n(\mu,\theta)}{\partial\mu\partial\beta\, \partial(\sigma^2)} , \tag{2.27}$$

$$X_5 = (\mu_A-\mu_T)(\hat{\beta}-\beta)(\hat{\sigma}^2-\sigma^2)^2 \frac{\partial^4 L_n(\mu,\sigma_x^2,\beta,\sigma_2^2)}{\partial\mu\, \partial\beta\, \partial(\sigma_2^2)^2}, \tag{2.28}$$

$$X_6 = (\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)(\hat{\sigma}^2-\sigma^2)\ \frac{\partial^3 L_n(\mu,\theta)}{\partial\mu^2\partial(\sigma^2)}\ , \tag{2.29}$$

and

$$X_7 = (\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)(\hat{\sigma}^2-\sigma^2)^2\ \frac{\partial^4 L_n(\mu,\sigma_x^2,\beta,\sigma_3^2)}{\partial\mu^2\ \partial(\sigma_3^2)^2}\ . \tag{2.30}$$

Here, the notation in 2.23 means that ΣX_i converges in probability to zero. The definition of convergence in probability to zero is the following [c.f. Mann and Wald (18)]:

Y_N converges in probability to zero if for any $\lambda > 0$,

$$\lim_{N\longrightarrow\infty} P(|Y_N| < \lambda) = 1 \quad .$$

Hence, when 2.23 holds, we can replace θ by $\hat{\theta}$, apply the s.p.r.t. procedure, and the resultant test is asymptotically equivalent to the s.p.r.t. with θ known.

We shall now state a theorem, a corollary, and two lemmas to be used in showing that 2.23 holds. But first, for completeness, we define precisely what we mean by the concept of order [c.f. Mann and Wald (18)].

For any sequence of positive numbers $\{f(N)\}$, we write $a_N = O[f(N)]$ (to be read, "a_N is of order $f(N)$") if there is a positive constant M, say, such that $|a_N| < Mf(N)$ for all N.

We now state the theorem, corollary, and two lemmas. The associated proofs are given in Section A of the appendix (Chapter VIII).

Theorem 1:

If

$$\text{(i)}\quad A_{ni} = O(n^{-1/2}), \qquad (i=1,2,\ldots,r),$$

$$\text{(ii)}\quad E(B_{ni}^2) = O(n^{-1}), \quad (i=1,2,\ldots,s),$$

and

$$\text{(iii)}\quad n^{-1}C_n \xrightarrow{(p)} 0,$$

then

$$Y_n = \left(\prod_{i=1}^{r} A_{ni}\right)\left(\prod_{i=1}^{s} B_{ni}\right) C_n \xrightarrow{(p)} 0 .$$

Corollary:

Theorem 1 holds for $r + s \geq 2$, i.e., for the cases $r = 0$, $s \geq 2$ and $s = 0$, $r \geq 2$.

Lemma 1:

If

$$\text{(i)}\quad D_n \xrightarrow{(p)} 0,$$

and

$$\text{(ii)}\quad E_n \xrightarrow{(p)} 0,$$

then

$$D_n E_n \xrightarrow{(p)} 0.$$

Lemma 2:

If

$$A_i \xrightarrow{(p)} 0, \quad (i=1,2,\ldots,k),$$

then

$$\prod_{i=1}^{k} A_i \xrightarrow{(p)} 0.$$

We now demonstrate that $X_i \xrightarrow{(p)} 0$, $(i=1,2,\ldots,7)$, and hence, by Lemma 2, that 2.23 holds.

From 2.10 and 2.24 we obtain

$$X_1 = -(\mu_A-\mu_T)(\hat{\beta}-\beta)\frac{\Sigma x}{\sigma^2}$$

In Section B of the appendix (Chapter VIII) the m.l. estimators of the parameters under consideration are given; in Section C they are corrected for bias, and it is shown that

$$V(\hat{\beta}) = O(n^{-1}) \tag{2.31}$$

and

$$V(\hat{\sigma}^2) = O(n^{-1}), \tag{2.32}$$

where the estimators are based upon sample size n.

Since

$$x_i \sim NI(x;0,\sigma_x^2),$$

it follows from Khintchine's theorem [c.f. Cramér (12, p. 253)] that

$$n^{-1}\,\Sigma x \xrightarrow{(p)} 0.$$

Recalling that $\mu_T-\mu$ and $\mu_A-\mu$ are taken to be of order $n^{-1/2}$, it follows that $\mu_A-\mu_T$ is of order $n^{-1/2}$. Also from 2.31 we have

$$E[(\hat{\beta}-\beta)^2] = V(\hat{\beta}) \doteq O(n^{-1}),$$

so that

$$X_1 \xrightarrow{(p)} \theta \tag{2.33}$$

by Theorem 1 where we let

$$\mu_A - \mu_T = A_{n1} ,$$

$$(\hat{\beta}-\beta) = B_{n1} ,$$

and

$$\Sigma x = C_n .$$

From 2.10 and 2.25 we obtain

$$X_2 = -(\mu_A-\mu_T)(\hat{\sigma}^2-\sigma^2) \frac{\Sigma(y-\mu-\beta x)}{\sigma^4}$$

$$= -(\mu_A-\mu_T)(\hat{\sigma}^2-\sigma^2)\frac{\Sigma(y-\mu)}{\sigma^4} + \beta(\mu_A-\mu_T)(\hat{\sigma}^2-\sigma^2)\frac{\Sigma x}{\sigma^4}$$

$$= X_{21} + X_{22} , \text{ say.}$$

We have

$$y_i \sim NI(y;\mu,\sigma_y^2)$$

so that

$$n^{-1}\, \Sigma(y-\mu) \xrightarrow{(p)} 0$$

by Khintchine's theorem. Also, from 2.32,

$$E(\hat{\sigma}^2-\sigma^2) = V(\hat{\sigma}^2) = O(n^{-1}).$$

Therefore

$$X_{21} \xrightarrow{(p)} 0$$

by Theorem 1 upon letting

$$(\mu_A-\mu_T) = A_{n1} ,$$

$$(\hat{\sigma}^2-\sigma^2) = B_{n1} ,$$

and

$$\Sigma(y-\mu) = C_n .$$

Also, taking

$$(\mu_A-\mu_T) = A_{n1} ,$$

$$(\hat{\sigma}^2-\sigma^2) = B_{n1} ,$$

and

$$\Sigma x = C_n$$

in Theorem 1, we have

$$X_{22} \xrightarrow{(p)} 0 ;$$

thus, by Lemma 2,

$$X_2 = X_{21} + X_{22} \xrightarrow{(p)} 0 . \tag{2.34}$$

We now consider X_3. Letting

$$(\mu_A-\mu_T) = A_{n1} ,$$

$$(\hat{\sigma}^2-\sigma^2) = B_{n1} = B_{n2} ,$$

and

$$\Sigma(y-\mu-\beta x) = C_n ,$$

2.10 and 2.26 give

$$X_3 = 2A_{n1}B_{n1}B_{n2}C_n\sigma_1^{-6}$$

$$= 2A_{n1}B_{n1}B_{n2}C_n\sigma^{-6} + 2A_{n1}B_{n1}B_{n2}C_n(\sigma_1^{-6}-\sigma^{-6})$$

$$= X_{31} + X_{32} \text{ , say.}$$

By Theorem 1,

$$X_{31} \xrightarrow{(p)} 0.$$

Now let

$$D_n = A_{n1}B_{n1}B_{n2}C_n$$

and

$$E_n = (\sigma_1^{-6}-\sigma^{-6}) \ .$$

Since z^{-3} is continuous in the neighborhood of $z = \sigma^2$, for any given $\Delta > 0$ there is a $\lambda > 0$ such that

$$|\sigma_1^2-\sigma^2| < \lambda \Rightarrow \quad \sigma_1^{-6}-\sigma^{-6}=|E_n| \ < \ \Delta^{1/2}$$

Further

$$P(|\hat{\sigma}^2-\sigma^2| < \lambda) \geq 1 - \frac{E[(\hat{\sigma}^2-\sigma^2)^2]}{\lambda^2}$$

by Tchebycheff's inequality [c.f. Cramér (12, p. 182)], and since

$$E[(\hat{\sigma}^2-\sigma^2)^2] = V(\hat{\sigma}^2) = O(n^{-1}) \ ,$$

it follows that

$$P(|\hat{\sigma}^2-\sigma^2| < \lambda) \longrightarrow 1 \ .$$

But since σ_1^2 lies between $\hat{\sigma}^2$ and σ^2 ,

$$|\sigma_1^2-\sigma^2| < |\hat{\sigma}^2-\sigma^2|$$

so that

$$|\hat{\sigma}^2-\sigma^2| < \lambda \Rightarrow |\sigma_1^2-\sigma^2| < \lambda \Rightarrow |E_n| < \Delta^{1/2} .$$

Therefore

$$P(|E_n| < \Delta^{1/2}) \geq P(|\hat{\sigma}^2-\sigma^2| < \lambda) \longrightarrow 1 .$$

In other words,

$$E_n \xrightarrow{(p)} 0$$

and since, by Theorem 1,

$$D_n \xrightarrow{(p)} 0 ,$$

Lemma 1 gives

$$X_{32} \xrightarrow{(p)} 0 .$$

Thus, it follows from Lemma 2 that

$$X_3 = X_{31} + X_{32} \xrightarrow{(p)} 0 . \tag{2.35}$$

We have from 2.10 and 2.27 that

$$X_4 = (\mu_A-\mu_T)(\hat{\beta}-\beta)(\hat{\sigma}^2-\sigma^2) \frac{\Sigma x}{\sigma^4}$$

and

$$X_4 \xrightarrow{(p)} 0 \tag{2.36}$$

from Theorem 1 by putting

$$(\mu_A-\mu_T) = A_{n1} ,$$

$$(\hat{\beta}-\beta) = B_{n1} ,$$

$$(\hat{\sigma}^2-\sigma^2) = B_{n2} ,$$

and

$$\Sigma x = C_n .$$

In considering X_5, we obtain from 2.10 and 2.28

$$X_5 = -2(\mu_A-\mu_T)(\hat{\beta}-\beta)(\hat{\sigma}^2-\sigma^2)^2(\Sigma x)\sigma^{-6}$$

$$-2(\mu_A-\mu_T)(\hat{\beta}-\beta)(\hat{\sigma}^2-\sigma^2)^2\Sigma x(\sigma_2^{-6}-\sigma^{-6})$$

$$= X_{51} - X_{52} , \text{ say.}$$

In Theorem 1, let

$$(\mu_A-\mu_T) = A_{n1} ,$$

$$(\hat{\beta}-\beta) = B_{n1} ,$$

$$(\hat{\sigma}^2-\sigma^2) = B_{n2} = B_{n3} ,$$

and

$$\Sigma x = C_n$$

so that

$$X_{51} \xrightarrow{(p)} 0 .$$

In Lemma 1, let

$$(\mu_A-\mu_T)(\hat{\beta}-\beta)(\hat{\sigma}^2-\sigma^2)\Sigma x = D_n$$

and

$$(\sigma_2^{-6}-\sigma^{-6}) = E_n$$

and we have

$$X_{52} \xrightarrow{(p)} 0 .$$

Therefore, by Lemma 2,

$$X_5 = X_{51} + X_{52} \xrightarrow{(p)} 0 . \qquad (2.37)$$

Referring to 2.10 and 2.29 we have

$$X_6 = (\mu_A - \mu_T)(\mu_A + \mu_T - 2\mu)(\hat{\sigma}^2 - \sigma^2)\,\frac{n}{\sigma^4} .$$

Since

$$|\mu_A + \mu_T - 2\mu| \le |\mu_A - \mu| + |\mu_T - \mu| ,$$

we have

$$(\mu_A + \mu_T - 2\mu) = O(n^{-1/2}) .$$

Further, we have that

$$(\hat{\sigma}^2 - \sigma^2) \xrightarrow{(p)} 0 .$$

Hence by letting

$$(\mu_A - \mu_T) = A_{n1} ,$$

$$(\mu_A + \mu_T - 2\mu) = A_{n2} ,$$

and

$$n(\hat{\sigma}^2 - \sigma^2) = C_n ,$$

application of the corollary to Theorem 1 gives

$$X_6 \xrightarrow{(p)} 0 . \qquad (2.38)$$

Finally, 2.10 and 2.30 give

$$X_7 = -2n(\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)(\hat{\sigma}^2-\sigma^2)^2\sigma^{-6}$$

$$-2n(\mu_A-\mu_T)(\mu_A+\mu_T-2\mu)(\hat{\sigma}^2-\sigma^2)^2(\sigma_3^{-6}-\sigma^{-6})$$

$$= X_{71} + X_{72} \text{ , say.}$$

By letting

$$(\mu_A-\mu_T) = A_{n1} ,$$

$$(\mu_A+\mu_T-2\mu) = A_{n2} ,$$

$$(\hat{\sigma}^2-\sigma^2) = B_{n1} ,$$

and

$$n(\hat{\sigma}^2-\sigma^2) = C_n ,$$

Theorem 1 gives

$$X_{71} \xrightarrow{(p)} 0 .$$

And by taking

$$n(\mu_A-\mu_T)(\mu_A+\mu_T)(\hat{\sigma}^2-\sigma^2) = D_n$$

and

$$(\sigma_3^{-6}-\sigma^{-6}) = E_n ,$$

Lemma 1 gives

$$X_{72} \xrightarrow{(p)} 0 .$$

Therefore

$$X_7 = X_{71} + X_{72} \xrightarrow{(p)} 0 \tag{2.39}$$

by Lemma 2.

Equations 2.33 — 2.39 give that

$$X_i \xrightarrow{(p)} 0 \quad (i=1,2,\ldots,7)$$

so we have obtained by Lemma 2 that 2.23 holds, i.e.

$$\sum_{i=1}^{7} X_i \xrightarrow{(p)} 0 \ .$$

Therefore, where we treat $\mu_A-\mu$ and $\mu_T-\mu$ as being of order $n^{-1/2}$, the Wald-type test of H_T against H_A, in which θ is replaced by $\hat{\theta}$, is asymptotically equivalent to the s.p.r.t. with θ known. Thus, the large sample test procedure is as follows:

At the n-th stage compute

$$\log P_n = L_n(\mu_A,\hat{\theta}) - L_n(\mu_T,\hat{\theta})$$

$$= \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} [(y_i-\mu_T-\hat{\beta}x_i)^2-(y_i-\mu_A-\hat{\beta}x_i)^2]$$

$$= \frac{n(\mu_A-\mu_T)}{\hat{\sigma}^2} (\bar{y}-\hat{\beta}\bar{x}- \frac{\mu_A+\mu_T}{2}) \ , \tag{2.40}$$

where $\hat{\sigma}^2$, $\hat{\beta}$, $\bar{y}$, and $\bar{x}$ — based on sample size n — are given by

$$\bar{y} = \frac{1}{n} \Sigma y \ , \tag{2.41}$$

$$\bar{x} = \frac{1}{n} \Sigma x \ , \tag{2.42}$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} , \tag{2.43}$$

and

$$\hat{\sigma}^2_{\text{(unbiased)}} = \frac{1}{n-2} \left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) , \tag{2.44}$$

where, for example,

$$S_{xx} = \Sigma(x-\bar{x})(x-\bar{x}) = \Sigma(x-\bar{x})^2 .$$

Then, if

$$\log \frac{\alpha_2}{1-\alpha_1} < \log P_n < \log \frac{1-\alpha_2}{\alpha_1} , \tag{2.45}$$

another observation is taken. If

$$\log P_n \geq \log \frac{1-\alpha_2}{\alpha_1} , \tag{2.46}$$

the process is terminated by accepting $H_A: \mu = \mu_A$. If

$$\log P_n \leq \log \frac{\alpha_2}{1-\alpha_1} , \tag{2.47}$$

the process is terminated by accepting $H_T: \mu = \mu_T$.

We now consider the rationale for treating $\mu_A - \mu$ and $\mu_T - \mu$ (and thus also $\mu_A - \mu_T$) as being of order $n^{-1/2}$. Many of the following remarks derive from private correspondence in 1964 with D. R. Cox, Birkbeck College, University of London, in which he explained in somewhat greater detail the approach used in his paper [Cox (10)].

Being a large sample approach, this scheme can be justified only in terms of a limiting operation in which the sample size required tends to infinity. Now if we fix the base hypotheses μ_A and μ_T, and consider a series of plans in which the sample sizes become large, the probabilities of error would become so minute that casual inspection of the observations would be adequate. So this approach does not lead to a very meaningful limiting operation. On the other hand, by separating μ_T and μ_A by an amount of order $n^{-1/2}$, we are considering a series of plans in each of which discrimination is possible with error probabilities roughly constant and which could be chosen to be of the magnitude used in applications.

The asymptotic result is of value, however, only if it gives a good guide to the behavior for the necessarily finite sample sizes to be encountered in practice. Of course this should be checked by comparing large sample and exact theory. However, the exact theory appears to be rather intractable in the present situation. It is hoped that the numerical investigation to be presented in Chapter IV will be helpful in making this comparison and in studying properties of the large sample test.

C. A Two-Sided Large Sample SPRT

In this section we shall describe the mechanics for conducting the large sample s.p.r.t. of the two-sided analogue of formulation 1. In this situation we wish to test

$H_T:\mu=\mu_T$ against $H_A:\mu=\pm\mu_A$, where for simplicity, we let $\pm\mu_A$ denote $\mu_T\pm d$ (d being a specified constant). We denote by H_A^+ and H_A^- the hypotheses that $\mu=\pm\mu_A$ and $\mu=-\mu_A$, respectively.

The procedure adopted herein is that developed by Armitage (4). The test of H_T against H_A^+ and H_A^- is a special case of a more general procedure due to Sobel and Wald (25) for choosing one of three hypotheses. The procedure involves conducting two large sample s.p.r.t.'s, one to test H_T against H_A^- and the other to test H_T against H_A^+. We shall denote these tests as T^- and T^+, respectively. Equation 2.40 gives the computational form for log P_n^- and log P_n^+, except, of course, that $-\mu_A$ is substituted for μ_A in computing log P_n^-.

To conduct the two sided test, we simultaneously carry out tests T^- and T^+ at each stage until one of the following events occur:

(i) One test terminates before the other.

(ii) Both tests terminate at the same stage.

If (i) occurs, the terminating test is no longer computed, and the other test is continued until it decisions. The decision rule is as follows:

(iii) If T^- accepts H_A^- and T^+ accepts H_T, we accept H_A^-.

(iv) If T^- accepts H_T and T^+ accepts H_T, we accept H_T.

(v) If T^- accepts H_T and T^+ accepts H_A^+, we accept H_A^+.

Sovel and Wald (25) show that the event "T^- accepts H_A^- and T^+ accepts H_A^+" cannot occur where α_1,α_2 in the limits for tests T^- and T^+ are the same.

The limits for both tests T^-, T^+ are taken to be

$\left(\log \frac{\alpha_2}{1-\alpha_1}, \log \frac{1-\alpha_2}{\alpha_1}\right)$ for

(vi) $P(\text{rejecting } H_T \mid H_T) = 2\alpha_1$,

(vii) $P(\text{accepting } H_T \mid H_A^-) = \alpha_2$,

and

(viii) $P(\text{accepting } H_T \mid H_A^+) = \alpha_2$.

III. WEIGHT-FUNCTION TESTS UTILIZING CONCOMITANT INFORMATION

A. Wald's Weight Functions

A brief outline of the weight-function approach to sequential testing of composite hypotheses as developed by Wald (28, pp. 80-83) is given in this section.

Let H_Δ be the composite hypothesis that the parameter point θ lies in a subset Δ of the parameter space. We divide the parameter space into three mutually exclusive zones: the zones of preference for acceptance Δ_a, the zone of preference for rejection Δ_r, and the zone of indifference.

In general, the probability of rejecting H_Δ when H_Δ is true will vary with the parameter point in Δ for any test procedure. Let $\alpha_1(\theta)$ denote the probability of an error of this type when θ is the parameter point. Similarly for any point θ lying outside Δ, let $\alpha_2(\theta)$ denote the probability of accepting H_Δ when H_Δ is false.

Ideally we wish to find a test procedure such that $\alpha_1(\theta)$ is less than or equal to a preassigned value α_1 for all θ in Δ_a, and such that $\alpha_2(\theta)$ will not exceed a preassigned value α_2 for all θ in Δ_r. But first we consider what shall be referred to herein as the modified problem. Let $w_a(\theta)$ and $w_r(\theta)$ by non-negative weight functions such that

$$\int_{\Delta_a} w_a(\theta)d\theta = 1 = \int_{\Delta_r} w_r(\theta)d\theta \tag{3.1}$$

The modified problem is to construct a test such that the weighted averages of the two types of error probabilities are equal to their preassigned values, i.e. such that

$$\int_{\Delta_a} \alpha_1(\theta) w_a(\theta) d\theta = \alpha_1$$

and (3.2)

$$\int_{\Delta_r} \alpha_2(\theta) w_r(\theta) d\theta = \alpha_2$$

Wald showed that a solution to the modified problem is given by the s.p.r.t. with test statistic given by

$$p_n^* = \frac{\int_{\Delta_r} w_r(\theta) \prod_{i=1}^{n} f(x_i;\theta)\, d\theta}{\int_{\Delta_a} w_a(\theta) \prod_{i=1}^{n} f(x_i;\theta)\, d\theta}, \tag{3.3}$$

where $f(x;\theta)$ is the p.d.f. of the population under investigation. Thus, in effect, we are actually testing the hypothesis that the p.d.f. of $x_1, x_2, \ldots, n_n$ is given by the denominator of 3.3 against the hypothesis that the numerator of 3.3 is the p.d.f. of $x_1, x_2, \ldots, x_n$. Wald proposed that this test will satisfy 3.2, and that for practical purposes, we may use

$$D = \frac{\alpha_2}{1-\alpha_1}$$

and

$$C = \frac{1-\alpha_2}{\alpha_1}$$

as limits for conducting the test.

Originally we sought a test procedure such that

$$\alpha_1(\theta) \leq \alpha_1, \quad \text{for all } \theta \text{ in } \Delta_a,$$

and (3.4)

$$\alpha_2(\theta) \leq \alpha_2, \quad \text{for all } \theta \text{ in } \Delta_r .$$

In addressing himself to this problem, Wald (28, p. 82) considers only s.p.r.t.'s given by 3.3, where $w_a(\theta)$ and $w_r(\theta)$ are any weight functions satifying 3.1. Denote by G the class of all such tests corresponding to all possible weight functions. A test in G is uniquely determined by the choice of $w_a(\theta)$ and $w_r(\theta)$ and the constants C, D. Therefore the maximum of $\alpha_1(\theta)$ over θ in Δ_a and the maximum of $\alpha_2(\theta)$ over θ in Δ_r are determined uniquely by $C,D,w_a(\theta)$, and $w_r(\theta)$. We let $\bar{\alpha}_1[C,D,w_a(\theta),w_r(\theta)]$ and $\bar{\alpha}_2[C,D,w_a(\theta),w_r(\theta)]$ denote these maxima respectively. Thus we have that for given values of C,D, the weight functions $w_a(\theta)$, $w_r(\theta)$ may be regarded as optimum if they simultaneously minimize $\bar{\alpha}_1[C,D,w_a(\theta),w_r(\theta)]$ and $\bar{\alpha}_2[C,D,w_a(\theta),w_r(\theta)]$. While no general method is available for constructing such weight functions $w_a(\theta)$ and $w_r(\theta)$, Wald (28, pp. 203, 204) does consider one class of cases for which optimum weight functions may be determined. Assume that the boundary of Δ_r is a surface S_r, and that one can find two weight functions $v_a(\theta)$ and $v_r(\theta)$ such that

$$\int_{\Delta_a} v_a(\theta)d\theta = 1 = \int_{S_r} v_r(\theta)dS_r \quad . \tag{3.5}$$

If the s.p.r.t. based on

$$P_n^* = \frac{\int_{S_r} v_r(\theta) \prod_{i=1}^{n} f(x_i;\theta)\, dS_r}{\int_{\Delta_a} v_a(\theta) \prod_{i=1}^{n} f(x_i;\theta)\, d\theta} \tag{3.6}$$

satisfies (i) $\alpha_1(\theta)$ is constant in Δ_a, (ii) $\alpha_2(\theta)$ is constant over S_r, and (iii) for any interior point θ of Δ_r, $\alpha_2(\theta)$ does not exceed the constant value of $\alpha_2(\theta)$ on S_r, then $v_a(\theta)$ and $v_r(\theta)$ are optimum in the sense described above. Further, Wald (28, pp. 204-207) demonstrates the existence of optimum weight functions for the case in which one wishes to test the hypothesis that the mean θ of the normal variable with unknown variance σ^2 is $\theta = \theta_0$ against the hypothesis that $|\theta - \theta_0| \geq \gamma\sigma$.

The weight-function approach is used in the remainder of this chapter to develop sequential test procedures which make use of concomitant information.

B. A Weight-Function Test, Formulation 1

Formulation 1 involves testing $H_T{:}\mu=\mu_T$ against $H_A{:}\mu=\mu_A$. For any positive constant c let the weight functions w_a and w_r be given as

$$w_{a,c}(\sigma,\beta) = \frac{1}{2c^2} \text{ if } 0<\sigma\leq c,\ -c\leq\beta\leq c, \text{ and } \mu=\mu_T,$$

$$= 0 \text{ otherwise;}$$

$$w_{r,c}(\sigma,\beta) = \frac{1}{2c^2} \text{ if } 0<\sigma\leq c,\ -c\leq\beta\leq c, \text{ and } \mu=\mu_A,$$

$$= 0 \text{ otherwise.}$$

Then, following Wald, a solution of the modified problem as described in the previous section (see 3.2) is given by the s.p.r.t. with test statistic

$$P_n^* = \frac{\displaystyle\int_0^c \int_{-c}^c \frac{1}{2c^2} \prod_{i=1}^n f(x_i,y_i;\mu_A,\beta,\sigma)\,d\beta\,d\sigma}{\displaystyle\int_0^c \int_{-c}^c \frac{1}{2c^2} \prod_{i=1}^n f(x_i,y_i;\mu_T,\beta,\sigma)\,d\beta\,d\sigma}$$

$$= \frac{\displaystyle\int_0^c \int_{-c}^c \prod_{i=1}^n f(x_i,y_i;\mu_A,\beta,\sigma)\,d\beta\,d\sigma}{\displaystyle\int_0^c \int_{-c}^c \prod_{i=1}^n f(x_i,y_i;\mu_A,\beta,\sigma)\,d\beta\,d\sigma}, \tag{3.7}$$

where, from our distributional assumptions of Section B, Chapter I, we have

$$f(x,y;\mu,\beta,\sigma) = g(x;0,\sigma_x^2)h(y \mid x;\mu+\beta x,\sigma^2)$$

$$= \frac{1}{2\pi\sigma_x\sigma} \exp[-\frac{x^2}{2\sigma_x^2} - \frac{(y-\mu-\beta x)^2}{2\sigma^2}]. \quad (3.8)$$

Therefore, when we consider the limiting case where $c \longrightarrow \infty$ (so that the ranges of β and σ have only their natural restrictions) as did Wald (28, p. 205), 3.7 becomes

$$P_n^* = \frac{\int_0^\infty \int_{-\infty}^\infty \sigma^{-n} \exp[-\frac{1}{2\sigma^2} \Sigma(y-\mu_A-\beta x)^2] d\beta d\sigma}{\int_0^\infty \int_{-\infty}^\infty \sigma^{-n} \exp[-\frac{1}{2\sigma^2} \Sigma(y-\mu_T-\beta x)^2] d\beta d\sigma}, \quad (3.9)$$

provided the integrals exist.

Upon completing the square in β and integrating with respect to β, we obtain

$$\int_0^\infty \int_{-\infty}^\infty \sigma^{-n} \exp[-\frac{1}{2\sigma^2} \Sigma(y-\mu-\beta x)^2] d\beta d\sigma$$

$$= \int_0^\infty (\frac{2\pi}{\Sigma x^2})^{1/2} \sigma^{1-n} \exp\left[\frac{[\Sigma(y-\mu)x]^2}{2\sigma^2 \Sigma x^2} - \frac{\Sigma(y-\mu)^2}{2\sigma^2}\right] d\sigma , \quad (3.10)$$

so that

$$P_n^* = \frac{\displaystyle\int_0^\infty \sigma^{1-n} \exp\left[\frac{[\Sigma(y-\mu_A)x]^2}{2\sigma^2\Sigma x^2} - \frac{\Sigma(y-\mu_A)^2}{2\sigma^2}\right]d\sigma}{\displaystyle\int_0^\infty \sigma^{1-n} \exp\left[\frac{[\Sigma(y-\mu_T)x]^2}{2\sigma^2\Sigma x^2} - \frac{\Sigma(y-\mu_T)^2}{2\sigma^2}\right]d\sigma}$$

$$= \frac{\displaystyle\int_0^\infty \sigma^{1-n} \exp\left[\sigma^{-2}[-A(\mu_A)]\right] d\sigma}{\displaystyle\int_0^\infty \sigma^{1-n} \exp\left[\sigma^{-2}[-A(\mu_T)]\right] d\sigma}, \qquad (3.11)$$

where

$$A(\mu) = \frac{\Sigma x^2\Sigma(y-\mu)^2 - [\Sigma(y-\mu)x]^2}{2\Sigma x^2} \qquad (3.12)$$

The integrals in 3.11 may be evaluated [c.f. Bateman (7, p. 313)] as

$$\int_0^\infty \sigma^{1-n} \exp\left[\sigma^{-2}[-A(\mu)]\right]d\sigma = 2^{-1}[A(\mu)]^{\frac{2-n}{2}}\,\Gamma\left(\frac{n-2}{2}\right), \qquad (3.13)$$

so that 3.11 becomes

$$P_n^* = \left[\frac{A(\mu_A)}{A(\mu_T)}\right]^{\frac{2-n}{2}}$$

$$= \left[\frac{\Sigma_x{}^2\Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{\Sigma x^2\Sigma(y-\mu_A)^2 - [\Sigma(y-\mu_A)x]^2}\right]^{\frac{n-2}{2}} . \qquad (3.14)$$

The validity of 3.13 is subject to the requirements that $n > 2$ and that

$$\Sigma x^2 \Sigma (y-\mu_T)^2 > [\Sigma(y-\mu_T)x]^2 \quad .$$

By Cauchy's inequality [c.f. Widder (29, p. 313)], we have

$$\Sigma x^2 \Sigma (y-\mu_T)^2 \geq [\Sigma(y-\mu_T)x]^2 \; .$$

Equality holds, excluding the trivial cases when $x_i = 0$ (all i) and/or $y_i - \mu_T = 0$ (all i), if and only if $x_i = k(y_i - \mu_T)$, k being a constant. Hence equality holds if and only if x and y have correlation 1, and this not a case of practical concern.

The s.p.r.t. is executed by computing the statistic P_n^* given in 3.14 and comparing it with accept-reject limits taken in practice to be $\left(\frac{\alpha_2}{1-\alpha_1}, \frac{1-\alpha_2}{\alpha_1}\right)$. Hence, at the n-th stage we compute P_n^*. If $P_n^* \leq \frac{\alpha_2}{1-\alpha_1}$, the process is terminated and we accept H_T^*. If $\frac{1-\alpha_2}{\alpha_1} \leq P_n^*$, the process is terminated by accepting H_A^*. If $\frac{\alpha_2}{1-\alpha_1} < P_n^* < \frac{1-\alpha_2}{\alpha_1}$, another observation is taken and we compute P_{n+1}^*, etc. Here H_T^* denotes the hypothesis that the p.d.f. of $(y_1|x_1)$, $(y_2|x_2)$, $\ldots, (y_n|x_n)$ is given by the denominator of 3.9; H_A^* denotes the hypothesis that the p.d.f. of our sample is given by the numerator of 3.9.

C. A Weight-Function Test, Formulation 2

Under formulation 2 we wish to test $H_T: \mu=\mu_T$ against $H_A: \mu=\mu_T + \gamma\sigma$ where γ is a specified constant. Following Wald, we define weight functions as follows. For any $c > 0$, let

$$w_{a,c}(\sigma,\beta) = \frac{1}{2c^2} \quad \text{for } 0<\sigma\leq c,\ -c\leq\beta\leq c, \text{ and } \mu=\mu_T$$
$$= 0 \text{ otherwise,}$$

and

$$w_{r,c}(\sigma,\beta) = \frac{1}{2c^2} \quad \text{for } 0<\sigma\leq c,\ -c\leq\beta\leq c, \text{ and } \mu=\mu_T+\gamma\sigma,$$
$$= 0 \text{ otherwise.}$$

A solution to the modified problem (see 3.2) is then given by the s.p.r.t. for which the test statistic is

$$P_n^* = \frac{\displaystyle\int_0^c \int_{-c}^c \frac{1}{2c^2} \prod_{i=1}^n f(x_i, y_i; \mu_T + \gamma\sigma + \beta x, \sigma^2)\, d\beta\, d\sigma}{\displaystyle\int_0^c \int_{-c}^c \frac{1}{2c^2} \prod_{i=1}^n f(x_i, y_i; \mu_T + \beta x, \sigma^2)\, d\beta\, d\sigma} \cdot \quad (3.15)$$

Passing to the limit as $c \longrightarrow \infty$, 3.15 becomes

$$P_n^* = \frac{\displaystyle\int_0^\infty \int_{-\infty}^\infty \sigma^{-n} \exp\left[-\frac{\Sigma(y-\mu_T-\gamma\sigma-\beta x)^2}{2\sigma^2}\right] d\beta\, d\sigma}{\displaystyle\int_0^\infty \int_{-\infty}^\infty \sigma^{-n} \exp\left[-\frac{\Sigma(y-\mu_T-\beta x)^2}{2\sigma^2}\right] d\beta\, d\sigma} \cdot \quad (3.16)$$

From Equations 3.9-3.13, the denominator of 3.16 is

$$\int_0^\infty \int_{-\infty}^\infty \sigma^{-n} \exp\left[-\frac{\Sigma(y-\mu_T-\beta x)^2}{2\sigma^2}\right] d\beta d\sigma$$

$$= \frac{1}{2}\left(\frac{2\pi}{\Sigma x^2}\right)^{1/2}\left[\frac{\Sigma x^2\ \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{2\Sigma x^2}\right]^{\frac{2-n}{2}} \Gamma\left(\frac{n-2}{2}\right) . \tag{3.17}$$

Further, upon completing the square in β and integrating with respect to β, the numerator of 3.16 becomes

$$\int_0^\infty \int_{-\infty}^\infty \sigma^{-n} \exp\left[-\frac{\Sigma(y-\mu_T-\gamma\sigma-\beta x)^2}{2\sigma^2}\right] d\beta d\sigma$$

$$= \left(\frac{2\pi}{\Sigma x^2}\right)^{1/2} \int_0^\infty \sigma^{1-n} \exp\left[-a\sigma^{-2}-b\sigma^{-1}+c\right] d\sigma , \tag{3.18}$$

where

$$a = \frac{\Sigma x^2\ \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{2\Sigma x^2} , \tag{3.19}$$

$$b = \frac{\gamma}{\Sigma x^2}\left[\Sigma(y-\mu_T)x\ \Sigma x - \Sigma x^2\ \Sigma(y-\mu_T)\right] , \tag{3.20}$$

and

$$c = \frac{\gamma^2}{2\Sigma x^2}\left[(\Sigma x)^2 - n\Sigma x^2\right] . \tag{3.21}$$

We have [c.f. Bateman (7, p. 313)]

$$\left(\frac{2\pi}{\Sigma x^2}\right)^{1/2}\int_0^\infty \sigma^{1-n} \exp\,[-a\sigma^{-2} - b\sigma^{-1} + c]\, d\sigma$$

$$= \left(\frac{2\pi}{\Sigma x^2}\right)^{1/2} \exp(c)(2a)^{\frac{2-n}{2}}\,\Gamma(n-2)\,\exp\left(\frac{b^2}{8a}\right)D_{2-n}[b(2a)^{-1/2}] \qquad (3.22)$$

where the parabolic cylinder function $D_q(z)$ is [c.f. Bateman (7, p. 386)]

$$D_q(z) = 2^{\frac{q}{2}} \exp\left(-\frac{z^2}{4}\right) \left[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1-q}{2})} F\left(-\frac{q}{2},\frac{1}{2};\frac{z^2}{2}\right)\right.$$

$$\left. + \frac{z}{2^{1/2}}\,\frac{\Gamma(-\frac{1}{2})}{\Gamma(-\frac{q}{2})} F\left(\frac{1-q}{2}, \frac{3}{2}; \frac{z^2}{2}\right)\right] \quad . \qquad (3.23)$$

In 3.23, $F(f,g;z)$ denotes the confluent hypergeometric series [c.f. (26, p. vi)]. The validity of 3.22 is subject to the restrictions that $n > 2$ and that

$$\Sigma x^2\, \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2 > 0 \quad .$$

The latter restriction was discussed in Section B of this chapter and shown to hold for cases of practical interest.

From Equations 3.17 - 3.23 we can now write the test statistic as

$$P_n^* = \frac{\Gamma(n-2)\exp(c)}{2^{n-3}\Gamma(\frac{n-2}{2})}\Bigg[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{n-1}{2})} F(\frac{n-2}{2}, \frac{1}{2}, \frac{b^2}{4a})$$

$$+ \frac{b\,\Gamma(-\frac{1}{2})}{2a^{1/2}\Gamma(\frac{n-2}{2})} F(\frac{n-1}{2}, \frac{3}{2}, \frac{b^2}{4a})\Bigg] , \qquad (3.24)$$

where a, b, and c are given by 3.19, 3.20, and 3.21, respectively.

In (26), tables of $F(\frac{m}{2}, \frac{1}{2}; z)$ for $z=0(.01)0.10$ and $m=3(2)201$, of $\frac{F(\frac{m}{2}, \frac{1}{2}; z)}{\cosh[z(2n-1)]^{1/2}}$ for $z=0(.01)0.10$ and $m=43(2)201$, and of $\frac{\log F(\frac{m}{2}, \frac{1}{2}; z)}{(2mz)^{1/2}}$ for $0.10\leq z\leq 100$ and $m=3(2)201$ are given. To make use of these tables, we note that [c.f. (26, p. vi)]

$$zF(m+1,\frac{1}{2}+1;z) = \frac{1}{2}[F(m+1, \frac{1}{2};z) - F(m, \frac{1}{2};z)] . \qquad (3.25)$$

Thus, 3.24 may be written

$$P_n^* = \frac{\Gamma(n-2)\ \exp(c)}{2^{n-3}\,\Gamma(\frac{n-2}{2})}\left[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{n-1}{2})} F(\frac{n-2}{2}, \frac{1}{2}; \frac{b^2}{4a})\right.$$

$$\left. + \frac{a^{1/2}\,\Gamma(-\frac{1}{2})}{b\,\Gamma(\frac{n-2}{2})}\left[F(\frac{n-1}{2}, \frac{1}{2}; \frac{b^2}{4a}) - F(\frac{n-3}{2}, \frac{1}{2}; \frac{b^2}{4a})\right]\right] .$$

(3.26)

An equivalent representation of 3.26, obtained by using well known properties of the gamma function, is

$$P_n^* = e^c F(\frac{n-2}{2},\frac{1}{2};\frac{b^2}{4a}) - \frac{(n-3)!\,\pi^{1/2} e^c a^{1/2}}{2^{n-4}[(\frac{n-4}{2})!]^2 b}[F(\frac{n-1}{2},\frac{1}{2};\frac{b^2}{4a}) - F(\frac{n-3}{2},\frac{1}{2};\frac{b^2}{4a})]$$

for n even,

$$= e^c F(\frac{n-2}{2},\frac{1}{2};\frac{b^2}{4a}) - \frac{2^{n-2}[(\frac{n-3}{2})!]^2 e^c a^{1/2}}{(n-3)!\,\pi^{1/2} b}[F(\frac{n-1}{2},\frac{1}{2};\frac{b^2}{4a}) - F(\frac{n-3}{2},\frac{1}{2};\frac{b^2}{4a})]$$

for n odd. (3.27)

Here the quantities a, b, and c are, as given previously in 3.19-3.21,

$$a = \frac{\Sigma x^2\ \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{2\Sigma x^2} ,$$

$$b = \frac{\gamma}{\Sigma x^2}\ [\Sigma(y-\mu_T)x\ \Sigma x - \Sigma x^2\ \Sigma(y-\mu_T)],$$

and

$$c = \frac{\gamma^2}{2\Sigma x^2}\,[(\Sigma x)^2 - n\Sigma x^2].$$

Since $F(0, \frac{1}{2}; z)$ is undefined, we require $n > 3$ for the existence of P_n^* given by 3.26 and 3.27. A second requirement not previously discussed is that $b \neq 0$, which will hold in general. If, however, due to the essentially discrete nature of our observations in practice, we have $b = 0$ at the n-th stage, say, the test statistic is defined simply as

$$P_n^* = e^c \quad . \tag{3.28}$$

Subject to these conditions, the s.p.r.t. with test statistic given by 3.26 - or equivalently by 3.27 - with limits $(\frac{\alpha_2}{1-\alpha_1}, \frac{1-\alpha_2}{\alpha_1})$ provides a solution to the modified problem. The test procedure and its interpretation are then as described in the preceding section for formulation 1.

D. Two-Sided Weight-Function Tests

In the two-sided analogue of formulation 2 we wish to test $H_T: \mu=\mu_T$ against $H_A: |\mu-\mu_T| = \gamma\sigma$, where γ is a specified positive constant. For this we now define our weight functions in the following manner. For any $c > 0$, let

$$w_{a,c}(\sigma,\beta) = \frac{1}{2c^2} \text{ for } 0 < \sigma \leq c,\ -c \leq \beta \leq c, \text{ and } \mu=\mu_T\ ,$$

$$= 0 \text{ otherwise},$$

and let

$$w_{r,c}(\sigma,\beta) = \frac{1}{4c^2} \text{ for } 0<\sigma\leq c,\ -c\leq\beta\leq c, \text{ and } \mu=\mu_T\pm\gamma\sigma ,$$

$$= 0 \text{ otherwise} .$$

A solution for the modified problem is given by the s.p.r.t. with test statistic

$$P_n^* = \frac{\displaystyle\int_0^c \int_{-c}^c \frac{\sigma^{-n}}{4c^2}\left[\exp -\left[\frac{\Sigma(y-\mu_T-\gamma\sigma-\beta x)^2}{2\sigma^2}\right]+\exp\left[-\frac{\Sigma(y-\mu_T+\gamma\sigma-\beta x)^2}{2\sigma^2}\right]\right]d\beta d\sigma}{\displaystyle\int_0^c \int_{-c}^c \frac{\sigma^{-n}}{2c^2} \exp\left[-\frac{\Sigma(y-\mu_T-\beta x)^2}{2\sigma^2}\right] d\beta d\sigma} . \qquad (3.29)$$

Passing to the limit as $c \longrightarrow \infty$, we obtain

$$P_n^* = \frac{\displaystyle\frac{1}{2}\int_0^\infty\int_{-\infty}^\infty \sigma^{-n}\left[\exp\left[-\frac{\Sigma(y-\mu_T-\gamma\sigma-\beta x)^2}{2\sigma^2}\right]+\exp\left[-\frac{\Sigma(y-\mu_T+\gamma\sigma-\beta x)^2}{2\sigma^2}\right]\right]d\beta d\sigma}{\displaystyle\int_0^\infty\int_{-\infty}^\infty \sigma^{-n} \exp\left[-\frac{\Sigma(y-\mu_T-\beta x)^2}{2\sigma^2}\right] d\beta d\sigma} . \qquad (3.30)$$

The denominator of 3.30 is given by 3.17, and from Equations 3.18 - 3.22 we see that the numerator of 3.30

$$\tfrac{1}{2}\left(\frac{2\pi}{\Sigma x^2}\right)^{\frac{1}{2}} \exp\left(c+\frac{b^2}{8a}\right)(2a)^{\frac{2-n}{2}} \Gamma(n-2)\left[D_{2-n}[b(2a)^{-\frac{1}{2}}]+D_{2-n}[-b(2a)^{\frac{1}{2}}]\right] . \tag{3.31}$$

Further, 3.23 gives

$$D_{2-n}[b(2a)^{-1/2}]+D_{2-n}[-b(2a)^{-1/2}]=2^{\frac{2-n}{2}} \exp\left(-\frac{b^2}{8a}\right)\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{n-1}{2})} {}_2F\left(\frac{n-2}{2},\frac{1}{2};\frac{b^2}{4a}\right). \tag{3.32}$$

Hence, Equations 3.17, 3.30, 3.31, and 3.32 show that the test statistic is

$$P_n^* = \frac{e^c\,\Gamma(n-2)\,\Gamma(\frac{1}{2})\ F(\frac{n-2}{2},\frac{1}{2};\frac{b^2}{4a})}{2^{n-3}\,\Gamma(\frac{n-2}{2})\,\Gamma(\frac{n-1}{2})}$$

$$= e^c\ F\left(\frac{n-2}{2},\frac{1}{2},\frac{b^2}{4a}\right) , \tag{3.33}$$

where, as in 3.19 - 3.21,

$$a = \frac{\Sigma x^2\ \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{2\Sigma x^2} ,$$

$$b = \frac{\gamma[\Sigma x(y-\mu_T)\Sigma x - \Sigma x^2\ \Sigma(y-\mu_T)]}{\Sigma x^2} ,$$

and

$$c = \frac{\gamma^2[(\Sigma x)^2 - n\Sigma x^2]}{2\Sigma x^2} \quad .$$

Again we use the limits $(\frac{\alpha_2}{1-\alpha_1}, \frac{1-\alpha_2}{\alpha_1})$ in conducting the test for the modified problem. However, in this case it is possible to make more direct usage of the tables given in (26) and thus lighten the computational task if we compute $\log P_n^*$. Then the test statistic is

$$\log P_n^* = c + \log F(\tfrac{n-2}{2},\tfrac{1}{2};\tfrac{b^2}{4a}) \quad , \tag{3.34}$$

with limits $(\log \frac{\alpha_2}{1-\alpha_1}, \log \frac{1-\alpha_2}{\alpha_1})$. Since (26) gives tables of

$\frac{\log F(\frac{m}{2},\frac{1}{2};z)}{(2mz)^{1/2}}$, $[0.10 \leq z \leq 100;\ m = 3(2)201]$, the test in form 3.34 is more readily computed than in form 3.33 when $\frac{b^2}{4a} \geq 0.10$.

The two-sided analogue of formulation 1 is

$$H_T: \mu=\mu_T \text{ versus } H_A: \mu=\pm\mu_A \ ,$$

where for simplicity, as in Chapter II, we let $\pm\mu_A$ denote $\mu_T \pm d$ (d being a specified constant) when $\mu_T \neq 0$.

For $c > 0$ we define our weight functions as

$$w_{a,c}(\sigma,\beta) = \frac{1}{2c^2} \text{ for } 0<\sigma\leq c,\ -c\leq\beta\leq c, \text{ and } \mu=\mu_T \ ,$$

$$= 0 \text{ otherwise;}$$

$$w_{r,c}(\sigma,\beta) = \frac{1}{4c^2} \text{ for } 0<\sigma\leq c,\ -c\leq\beta\leq c, \text{ and } \mu=\pm\mu_A \ ,$$

$$= 0 \text{ otherwise.}$$

Proceeding in a manner directly analogous to that employed in Section B of this chapter, we obtain the following s.p.r.t. statistic for the modified problem (see 3.2):

$$P_n^* = \frac{\frac{1}{2}\int_0^\infty\int_{-\infty}^\infty \sigma^{-n}\left[\exp[-\frac{\Sigma(y-\mu_A-\beta x)^2}{2\sigma^2}]+\exp[-\frac{\Sigma(y+\mu_A-\beta x)^2}{2\sigma^2}]\right]d\beta d\sigma}{\int_0^\infty\int_{-\infty}^\infty \sigma^{-n}\exp[-\frac{\Sigma(y-\mu_T-\beta x)^2}{2\sigma^2}]\ d\beta d\sigma}$$

$$= \frac{[A(\mu_A)]^{\frac{2-n}{2}} + [A(-\mu_A)]^{\frac{2-n}{2}}}{2[A(\mu_T)]^{\frac{2-n}{2}}} \ , \qquad (3.35)$$

where

$$A(\mu) = \frac{\Sigma x^2\ \Sigma(y-\mu)^2 - [\Sigma(y-\mu)x]^2}{2\Sigma x^2} \ .$$

Hence, 3.35 may be written as

$$P_n^* = \frac{1}{2}\left[\frac{\Sigma x^2\, \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{\Sigma x^2\, \Sigma(y-\mu_A)^2 - [\Sigma(y-\mu_A)x]^2}\right]^{\frac{n-2}{2}}$$

$$+ \frac{1}{2}\left[\frac{\Sigma x^2\, \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{\Sigma x^2\, \Sigma(y+\mu_A)^2 - [\Sigma(y+\mu_A)x]^2}\right]^{\frac{n-2}{2}} \qquad (3.36)$$

and in practice we take the limits to be $(\frac{\alpha_2}{1-\alpha_1}, \frac{1-\alpha_2}{\alpha_1})$.

E. A Weight-Function Test, Formulation 1 Modified

An apparent difficulty associated with the weight-function test for formulation 1 is discussed in Chapter IV. At that point it will be seen that the practical implementation of the test requires choosing the base hypotheses μ_T and μ_A so that one of them is close to the true value μ. To avoid this difficulty we consider the following modification of formulation 1:

$$H_T\!:\mu\leq\mu_T;\ H_A\!:\mu\geq\mu_A;\ (\mu_T<0<\mu_A). \qquad (3.37)$$

It is noted that the condition $\mu_T<0<\mu_A$ does not limit the situations to which the test procedure has application.

If in practice the test of interest (with observations x,y) is $\mu \leq 0$ against $\mu \geq d$ (d being a specified positive constant), an equivalent test may be obtained by testing, for example, $\mu \leq -\frac{d}{2}$ against $\mu \geq \frac{d}{2}$ with observations $x, (y-\frac{d}{2})$.

We now develop the test procedure for the modification of formulation 1. For any $c > 1$, we define

$$w_{a,c}(\mu,\sigma,\beta) = -\frac{1}{2c^2(c-1)\mu_T}, \text{ for } 0<\sigma\leq c, -c\leq\beta\leq c, \text{ and } c\mu_T\leq\mu\leq\mu_T,$$

$$= 0, \text{ otherwise,}$$

and

$$w_{r,c}(\mu,\sigma,\beta) = \frac{1}{2c^2(c-1)\mu_A}, \text{ for } 0<\sigma\leq c, -c\leq\beta\leq c, \text{ and } \mu_A\leq\mu\leq c\mu_A,$$

$$= 0, \text{ otherwise.}$$

Then

$$P_n^* = -\frac{\mu_T}{\mu_A} \frac{\int_{\mu_A}^{c\mu_A}\int_0^c\int_{-c}^{c} \sigma^{-n} \exp[-\frac{\Sigma(y-\mu-\beta x)^2}{2\sigma^2}]d\beta d\sigma d\mu}{\int_{c\mu_T}^{\mu_T}\int_0^c\int_{-c}^{c} \sigma^{-n} \exp[-\frac{\Sigma(y-\mu-\beta x)^2}{2\sigma^2}]d\beta d\sigma d\mu}. \tag{3.38}$$

Passing to the limit as $c \longrightarrow \infty$, 3.38 becomes

$$P_n^* = -\frac{\mu_T}{\mu_A}\frac{\int_{\mu_A}^{\infty}\int_0^{\infty}\int_{-\infty}^{\infty}\sigma^{-n}\exp\left[-\frac{\Sigma(y-\mu-\beta x)^2}{2\sigma^2}\right]d\beta d\sigma d\mu}{\int_{-\infty}^{\mu_T}\int_0^{\infty}\int_{-\infty}^{\infty}\sigma^{-n}\exp\left[-\frac{\Sigma(y-\mu-\beta x)^2}{2\sigma^2}\right]d\beta d\sigma d\mu} \tag{3.39}$$

From 3.17 we obtain

$$P_n^* = -\frac{\mu_T}{\mu_A}\frac{\int_{\mu_A}^{\infty}[B(\mu)]^{\frac{2-n}{2}}\,d\mu}{\int_{-\infty}^{\mu_T}[B(\mu)]^{\frac{2-n}{2}}\,d\mu}, \tag{3.40}$$

where

$$B(\mu) = \Sigma x^2\,\Sigma(y-\mu)^2 - [\Sigma(y-\mu)x]^2 , \tag{3.41}$$

and subject to $n > 2$ and $B(\mu) > 0$. The requirement that $B(\mu) > 0$ was discussed in Section B of this chapter.

We may write $B(\mu)$ as

$$B(\mu) = h\mu^2 + g\mu + f, \tag{3.42}$$

where

$$f = \overset{n}{\Sigma}x^2\,\overset{n}{\Sigma}y^2 - (\overset{n}{\Sigma}xy)^2 , \tag{3.43}$$

$$g = 2(\overset{n}{\Sigma}x\,\overset{n}{\Sigma}xy - \overset{n}{\Sigma}x^2\,\Sigma y) , \tag{3.44}$$

and

$$h = n\overset{n}{\Sigma}x^2 - (\overset{n}{\Sigma}x)^2 \, . \tag{3.45}$$

Here we have, by Cauchy's inequality [c.f. Widder (29, p. 313)],

$$f = \Sigma x^2 \, \Sigma y^2 - (\Sigma xy)^2 \geq 0. \tag{3.46}$$

Equality holds in 3.46 only in uninteresting cases as pointed out in Section B of this chapter in conjunction with the requirement that $B(\mu)$ be greater than zero.

Also

$$h = n\Sigma x^2 - (\Sigma x)^2 = n\Sigma(x-\bar{x})^2 \geq 0 \, , \tag{3.47}$$

with equality holding only when x_i = constant for all $i=1,2,\ldots,n$. Hence, for practical purposes, we need to consider only the case for which strict inequality holds in 3.46 and 3.47.

Let us now define

$$q = 4fh - g^2 \tag{3.48}$$

and

$$k = \frac{4h}{q} \, . \tag{3.49}$$

Further, since $B(\mu) > 0$ its discriminant is less than zero, i.e.

$$-q = g^2 - 4fh < 0$$

so that

$$q > 0 \, . \tag{3.50}$$

We make use of the following standard integral forms:

For m a positive integer,

$$\int \frac{d\mu}{[B(\mu)]^{m+1}} = \frac{2h\mu+g}{mq[B(\mu)]^{m}} + \frac{2(2m-1)h}{mq} \int \frac{d\mu}{[B(\mu)]^{m}}, \quad (3.51)$$

$$\int \frac{d\mu}{B(\mu)} = 2q^{-1/2} \tan^{-1} [q^{-1/2}(2h\mu + g)] , \quad (3.52)$$

$$\int \frac{d\mu}{[B(\mu)]^{3/2}} = \frac{2(2h\mu+g)}{qB(\mu)} , \quad (3.53)$$

and, for $m \geq 2$,

$$\int \frac{d\mu}{[B(\mu)]^{m+1/2}} = \frac{2(2h\mu+g)}{(2m-1)q[B(\mu)]^{m-1/2}} + \frac{2k(m-1)}{2m-1} \int \frac{d\mu}{[B(\mu)]^{m-1/2}} . \quad (3.54)$$

We have now established the framework for investigating the improper integrals in 3.40. We write the indefinite integral

$$\int [B(\mu)]^{\frac{2-n}{2}} d\mu = \int \frac{d\mu}{[B(\mu)]^{\frac{n}{2}-1}} = I_n(\mu) . \quad (3.55)$$

Then from Equations 3.51, 3.52, and 3.55 we obtain the recursion relationship

$$I_n(\mu) = \frac{2(2h\mu+g)}{(n-4)q[B(\mu)]^{\frac{n-4}{2}}} + \frac{4h(n-5)}{q(n-4)} I'_{n-2}(\mu) \qquad (3.56)$$

for n even (n=6,8,10,...), where

$$I'_4 = 2q^{-1/2} \tan^{-1} [q^{-1/2}(2h\mu+g)] \ , \qquad (3.57)$$

and the prime (') appearing in 3.56 and 3.57 denotes that the quantities g, h, and q are based upon sample size n. Similarly Equations 3.53 - 3.55 give

$$I_n(\mu) = \frac{2(2h\mu+g)}{(n-4)q[B(\mu)]^{\frac{n-4}{2}}} + \frac{k(n-5)}{(n-4)} I'_{n-2}(\mu) \qquad (3.58)$$

for n odd (n=7,9,11,...) and

$$I'_5(\mu) = \frac{2(2h\mu+g)}{(n-4)q[B(\mu)]^{1/2}} \ . \qquad (3.59)$$

Since k = 4h/q, Equation 3.58 is the same as 3.56 and we write

$$I_n(\mu) = \frac{2(2h\mu+g)}{(n-4)q[B(\mu)]^{\frac{n-4}{2}}} + \frac{k(n-5)}{(n-4)} I'_{n-2}(\mu) \ , \qquad (3.60)$$

which holds for n odd or even (n=6,7,8,...).

We now consider limiting forms of $I_n(\mu)$ as μ tends to

$+\infty$ and as μ tends to $-\infty$. First we consider $I_4'(\mu)$ given in 3.57. Where $q^{-1/2}$ denotes the positive root of q, we have

$$\lim_{\mu\to\infty} I_4'(\mu) = 2q^{-1/2} \lim_{\mu\to\infty} \tan^{-1}[q^{-1/2}(2h\mu+g)]$$

$$= 2q^{-1/2}(\frac{\pi}{2} \pm 2r\pi) \ , \qquad (3.61)$$

where r=0,1,2,..., and we call $\pi/2$ the "principal value" of $\lim_{\mu\to\infty} \tan^{-1}\mu$. Further discussion of principal values follows shortly. Similarly,

$$\lim_{\mu\to-\infty} I_4'(\mu) = 2q^{-1/2}(-\frac{\pi}{2} \pm 2r\pi) \ . \qquad (3.62)$$

Where $h^{1/2}$ denotes the positive root of h, we obtain

$$\lim_{\mu\to\infty} I_5'(\mu) = \frac{2}{q} \lim_{\mu\to\infty} \frac{2h\mu+g}{[B(\mu)]^{1/2}}$$

$$= \frac{2}{q} \lim_{\mu\to\infty} \frac{\mu(2h+\frac{g}{\mu})}{\mu(h+\frac{g}{\mu}+\frac{f}{\mu^2})^{1/2}} \ ,$$

by 3.42,

$$= \frac{2}{q}\frac{(2h)}{h^{1/2}} \quad \frac{4h^{1/2}}{q} \ , \qquad (3.63)$$

and similarly,

$$\lim_{\mu \to -\infty} I_5'(\mu) = -\frac{4h^{1/2}}{q} . \tag{3.64}$$

From Equations 3.60 - 3.62 we obtain, for n even (n=6,8,10,...)

$$\lim_{\mu \to \infty} I_n(\mu) = \frac{2}{(n-4)q} \frac{\mu(2h+\frac{g}{\mu})}{\mu^{n-4}(h+\frac{g}{\mu}+\frac{f}{\mu^2})^{\frac{n-4}{2}}} + \frac{k(n-5)}{(n-4)} \lim_{\mu \to \infty} I_{n-2}'(\mu)$$

$$= 0 + \frac{k(n-5)}{(n-4)} \lim_{\mu \to \infty} I_{n-2}'(\mu)$$

$$= 2q^{-1/2}(\frac{\pi}{2} \pm 2r\pi) \prod_{(j=6,8,10,\ldots,n)} \frac{k(j-5)}{(j-4)}$$

$$= 2q^{-1/2}(\frac{\pi}{2} \pm 2r\pi) \frac{k^{\frac{n-4}{2}}[(n-4)!]}{2^{n-4}[(\frac{n-4}{2})!]^2} ,$$

and

$$\lim_{\mu \to -\infty} I_n(\mu) = 2q^{-1/2}(-\frac{\pi}{2} \pm 2r\pi) \frac{k^{\frac{n-4}{2}}[(n-4)!]}{2^{n-4}[(\frac{n-4}{2})!]^2} . \tag{3.65}$$

Similarly, from Equations 3.60, 3.63, and 3.64 we obtain for n odd (n=7,9,11,...)

$$\lim_{\mu \to \infty} I_n(\mu) = -\lim_{\mu \to -\infty} I_n(\mu)$$

$$= \frac{4h^{1/2}}{q} \prod_{(j=7,9,11,\ldots n)} \frac{k(j-5)}{(j-4)}$$

$$= \frac{4h^{1/2}}{q} \frac{k^{\frac{n-5}{2}} 2^{n-5}[(\frac{n-5}{2})!]^2}{(n-4)[(n-5)!]} . \qquad (3.66)$$

From 3.40 and 3.55 we obtain

$$P_n^* = - \frac{\mu_T[\lim_{\mu \to \infty} I_n(\mu) - I_n(\mu_A)]}{\mu_A[I_n(\mu_T) - \lim_{\mu \to -\infty} I_n(\mu)]} . \qquad (3.67)$$

Therefore 3.57, 3.61, 3.62 and 3.67 give

$$P_4^* = - \frac{\mu_T}{\mu_A} \left[\frac{2q^{-1/2}(\frac{\pi}{2} \pm 2r\pi) - 2q^{-1/2}\tan^{-1}[q^{-1/2}(2h\mu_A+g)]}{2q^{-1/2}\tan^{-1}[q^{-1/2}(2h\mu_T+g)]-2q^{-1/2}(-\frac{\pi}{2} \pm 2r\pi)} \right] . \qquad (3.68)$$

In the equation

$$\tan^{-1} w = z,$$

z has the form

$$z = \theta \pm 2r\pi \quad (r=0,1,2,\ldots) ,$$

where θ, the "principal value," may be taken as

$$-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2} .$$

Hence, where Θ_A and Θ_T denote principal values in

$$z_A = \Theta_A \pm 2r\pi = \tan^{-1}[q^{-1/2}(2h\mu_A+g)]$$

and

$$z_T = \Theta_T \pm 2r\pi = \tan^{-1}[q^{-1/2}(2h\mu_A+g)] \ ,$$

3.68 becomes

$$P_4^* = - \frac{\mu_T[2q^{-1/2}(\frac{\pi}{2} - \Theta_A)]}{\mu_A[2q^{-1/2}(\Theta_T + \frac{\pi}{2})]} \ . \tag{3.69}$$

Alternatively, by considering only principal values in 3.67, 3.68 may be written

$$P_4^* = - \frac{\mu_T}{\mu_A}\left[\frac{\pi-2\ \tan^{-1}[q^{-1/2}(2h\mu_A+g)]}{\pi+2\ \tan^{-1}[q^{-1/2}(2h\mu_T+g)]}\right] \ . \tag{3.70}$$

From Equations 3.59, 3.63, 3.64, and 3.67 we obtain

$$P_5^* = \frac{-\mu_T\left[\dfrac{4h^{1/2}}{q} - \dfrac{2(2h\mu_A+g)}{q[B(\mu_A)]^{1/2}}\right]}{\mu_A\left[\dfrac{2(2h\mu_T+g)}{q[B(\mu_T)]^{1/2}} + \dfrac{4h^{1/2}}{q}\right]}$$

$$= -\frac{\mu_T}{\mu_A}\left[\frac{B(\mu_T)}{B(\mu_A)}\right]^{1/2}\left[\frac{2[hB(\mu_A)]^{1/2} - 2h\mu_A - g}{2[hB(\mu_T)]^{1/2} + 2h\mu_T + g}\right] \qquad (3.71)$$

Where we again consider only principal values of $\tan^{-1}w$, for n=6,8,10,..., we obtain from 3.65 and 3.67

$$P_n^* = -\frac{\mu_T}{\mu_A}\left[\frac{\dfrac{\pi q^{-1/2}k^{\frac{n-4}{2}}[(n-4)!]}{2^{n-4}[(\frac{n-4}{2})!]^2} - I_n(\mu_A)}{\dfrac{\pi q^{-1/2}k^{\frac{n-4}{2}}[(n-4)!]}{2^{n-4}[(\frac{n-4}{2})!]^2} + I_n(\mu_T)}\right]$$

$$= -\frac{\mu_T}{\mu_A}\left[\frac{\pi q^{-1/2}k^{\frac{n-4}{2}}[(n-4)!] - 2^{n-4}[(\frac{n-4}{2})!]^2 I_n(\mu_A)}{\pi q^{-1/2}k^{\frac{n-4}{2}}[(n-4)!] + 2^{n-4}[(\frac{n-4}{2})!]^2 I_n(\mu_T)}\right].$$

(3.72)

For n=7,9,11,..., we obtain from 3.66 and 3.67

$$P_n^* = -\frac{\mu_T}{\mu_A}\,\frac{\dfrac{4h^{1/2}k^{\frac{n-5}{2}}\,2^{n-5}[(\frac{n-5}{2})!]^2}{q(n-4)[(n-5)!]} - I_n(\mu_A)}{\dfrac{4h^{1/2}k^{\frac{n-5}{2}}\,2^{n-5}[(\frac{n-5}{2})!]^2}{q(n-4)(n-5)!} + I_n(\mu_T)}$$

$$= -\frac{\mu_T}{}\left[\frac{2^{n-3}h^{1/2}k^{\frac{n-5}{2}}[(\frac{n-5}{2})!]^2 - q(n-4)[(n-5)!]I_n(\mu_A)}{2^{n-3}h^{1/2}k^{\frac{n-5}{2}}[(\frac{n-5}{2})!]^2 + q(n-4)[(n-5)!]I_n(\mu_T)}\right]. \tag{3.73}$$

Equations 3.70-3.73 give the test statistic for the s.p.r.t. which, from Wald (28, pp. 81,82), provides a solution to the modified problem discussed in Section A of this chapter (c.f. Equation 3.2), where the quantities g,h,q,k, $B(\mu)$, and $I_n(\mu)$ are given in 3.41 - 3.45, 3.48, 3.49 and 3.56. In practice the limits are taken to be $(\frac{\alpha_2}{1-\alpha_1}, \frac{1-\alpha_2}{\alpha_1})$.

In this chapter, sequential test procedures utilizing concomitant information have been developed using Wald's weight-function approach. It can be seen that the test statistics are involved functions of the observations not readily amenable to theoretical study of their properties. Some numerical results on these are, however, discussed in Chapter IV.

IV. NUMERICAL INVESTIGATIONS

A. A Conjecture About the Reduction in Average Sample Number Due to Use of Concomitant Information

Consider testing $H_T:\mu=\mu_T$ against $H_A:\mu=\mu_A$ with normal observations y where we want the probability of rejecting H_T, given H_T is true, to be α_1, and the probability of rejecting H_A, given H_A is true, to be α_2. When concomitant information is not used, let N_f denote the average sample number (a.s.n.) required to detect the difference $\mu_A-\mu_T$ as significant in fixed sample number experiments and N_s denote the corresponding a.s.n. in sequential experiments. Further, when concomitant information is used, let C_f and C_s denote the a.s.n.'s required to detect $\mu_A-\mu_T$ as being significant in fixed sample number and sequential experiments respectively. It is then conjectured that, approximately,

$$\frac{C_s}{N_s} = \frac{C_f}{N_f} . \tag{4.1}$$

Cochran (9) states that use of covariance analysis (in fixed sample number experiments) reduces the experimental error variance σ_y^2 to a value which is effectively about

$$\sigma_y^2(1-\rho^2)\left(\frac{f_e-1}{f_e-2}\right) , \tag{4.2}$$

where f_e is the number of error degrees of freedom, and the factor involving f_e is needed to allow for errors in the estimated regression coefficient. Further, N_f is proportional to σ_y^2; for example Davies (13, p. 32) gives

$$N_f = \frac{(u_{\alpha_1} + u_{\alpha_2})^2 \sigma_y^2}{(\mu_A - \mu_T)^2} , \tag{4.3}$$

where u_{α_i} is the deviate of the standard normal curve which cuts off a single-tail area α_i. Then, from 4.2 and 4.3 we take C_f to be proportional to $\sigma_y^2(1-\rho^2)$ and obtain, approximately,

$$\frac{C_f}{N_f} = 1-\rho^2 \tag{4.4}$$

so that our conjecture in 4.1 becomes

$$\frac{C_s}{N_s} = 1 - \rho^2 . \tag{4.5}$$

The conjectured a.s.n. reduction attributed to utilizing concomitant information in sequential experiments is therefore

$$N_s - C_s \cong \rho^2 N_s , \tag{4.6}$$

i.e. the reduction is proportional to the square of the correlation coefficient.

Numerical evidence bearing on this conjecture is presented in the following section.

B. Monte Carlo Trials on the Large-Sample Test

Data for a numerical study were generated on an IBM 7074 at Iowa State University by the following procedure. First, observations u were generated from the uniform distribution by the power residue method as described in (14). Then univariate normal observations z were obtained, each from m = 12 uniformly distributed observations, as

$$z = \frac{\sum_{i=1}^{m} u_i - mE(u_i)}{(\frac{m}{12})^{1/2}} ,$$

In this way, two independent sets of observations, z_{1i} and z_{2i}, were obtained by generating each set independently from N(0,1). Then, to obtain bivariate normal observations (x,y) with E(x) = E(y) = 0, $\sigma_x^2 = \sigma_y^2 = 1$, and correlation coefficient ρ, we let

$$y = z_1$$

and (4.7)

$$x = (1-\rho^2)^{1/2} z_2 + \rho z_1 .$$

In the subsequent Monte Carlo trials it was required that E(y) take values other than zero, and this was accomplished by adding a constant to each y observation.

Numerical studies were conducted on the following test procedures: the large-sample test with covariance developed in Chapter II, which we denote as L_1: the corresponding large-sample test without covariance, L_2; the s.p.r.t. with covariance, S_1; the s.p.r.t. without covariance, S_2.

The logs of the n-th stage test statistics for the above tests are as follows:

L_1 (see Equation 2.40):

$$\frac{n(\mu_A-\mu_T)}{\hat{\sigma}^2}\left(\bar{y}-\hat{\beta}\bar{x}-\frac{\mu_A+\mu_T}{2}\right), \tag{4.8}$$

where

$$\hat{\sigma}^2 = \frac{1}{n-2}\left[\sum^{n}(y-\bar{y})^2 - \frac{[\Sigma(y-\bar{y})(x-\bar{x})]^2}{\sum^{n}(x-\bar{x})^2}\right]$$

$$= \frac{1}{n-2}\left(S_{yy}-\frac{S_{xy}^2}{S_{xx}}\right),$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}},$$

and $\bar{x}$, $\bar{y}$ are based on sample size $n(n=3,4,\ldots)$;

L_2:

$$\frac{n(\mu_A-\mu_T)}{\hat{\sigma}_y^2}\left(\bar{y}-\frac{\mu_A+\mu_T}{2}\right), \tag{4.9}$$

where

$$\hat{\sigma}_y^2 = \frac{S_{yy}}{n-1},$$

and $\bar{y}$ is based on sample size $n(n=2,3,\ldots)$;

S_1:

$$\frac{n(\mu_A-\mu_T)}{\sigma^2}(\bar{y} - \beta\bar{x} - \frac{\mu_A+\mu_T}{2}) \ , \tag{4.10}$$

where

$$\sigma^2 = \sigma_y^2(1-\rho^2) \ ,$$

and

$$\beta = \rho\frac{\sigma_y}{\sigma_x} \ ;$$

S_2 [see Wald (28, pp. 118-120)]:

$$\frac{n(\mu_A-\mu_T)}{\sigma_y^2}(\bar{y} - \frac{\mu_A+\mu_T}{2}) \ . \tag{4.11}$$

For all of these tests, the boundaries are taken to be $\log \frac{\alpha_2}{1-\alpha_1}$ and $\log \frac{1-\alpha_2}{\alpha_1}$; the combinations of (α_1, α_2) used in the numerical studies were (0.05, 0.05), (0.05, 0.01), and (0.01, 0.05), and we denote the corresponding boundaries by B_1, B_2, and B_3 respectively.

In all trials we set

$$\mu_x = 0 \ ,$$

$$\sigma_x = 1 \ ,$$

$$\sigma_y = 1 \ ,$$

and

$$\mu_T = 0 \ . \tag{4.12}$$

In order to obtain diverse testing conditions and to check error rate realizations, μ_A and $\mu_y = \mu$ were given different values in different series of trials. Also, σ was altered by letting ρ take values 0.60, 0.75, and 0.90. In addition this provided a basis for investigating the a.s.n. conjecture discussed in Section A of this chapter.

In early numerical studies of the large-sample tests L_1 and L_2 it was found that the error rates on the base hypotheses were far in excess of the nominal prespecified values α_1, α_2. For example — from a representative series of 250 trials with $\mu = 0$, $\mu_T = 0$, $\mu_A = 0.4$, and $\rho = 0.75$ and from another series of 250 trials with $\mu = 0.4$, $\mu_T = 0$, $\mu_A = 0.4$, and $\rho = 0.75$ — the results for tests L_1 and L_2 were as follows, where we let α_1', α_2' denote observed error rates corresponding to the specified theoretical error rates α_1, α_2:

	L_1	L_2
α_1, α_2	α_1', α_2'	α_1', α_2'
0.05, 0.05	0.264, 0.204	0.156, 0.156
0.05, 0.01	0.264, 0.168	0.164, 0.080
0.01, 0.05	0.196, 0.204	0.145, 0.160

The corresponding results for tests S_1 and S_2 were:

α_1, α_2	S_1: α_1', α_2'	S_2: α_1', α_2'
0.05, 0.05	0.032, 0.048	0.036, 0.052
0.05, 0.01	0.036, 0.004	0.040, 0.000
0.01, 0.05	0.008, 0.048	0.004, 0.048

It appeared as though this difficulty obtained due to the large variability of the associated test statistics, in particular due to the variability of the variance estimators, at early stages when few degrees of freedom were available for estimation. The minimum sample numbers for test L_1 and L_2 are n = 3 and n = 2 respectively, at which point only one degree of freedom is available for variance estimation. In the above series of 250 trials with $(\mu, \mu_T, \mu_A, \rho) = (0, 0, 0.4, 0.75)$, for example, test L_1 decisioned 116 times on the third stage with boundary B_1 (corresponding to $\alpha_1, \alpha_2 = 0.05, 0.05$) and had a.s.n. = 8.40; test L_2 decisioned 70 times on the second stage with boundary B_1 and had a.s.n. = 22.34. The corresponding a.s.n.'s for S_1 and S_2 were 16.64 and 39.32 respectively.

To obviate this difficulty associated with tests L_1 and L_2, an empirical censoring scheme was introduced in which supposedly premature decisions were ignored and sampling was continued until certain minimum stage numbers N_{min} were attained after which a decision was recognized.

The computer program used for the numerical studies made provision for considering seven different levels of censoring N_{min} during each series of trials on a particular parametric combination (μ,μ_T,μ_A,ρ), but it was required that the N_{min} be specified at the beginning of each series. Further, to study the type I error α_1 it was necessary to use parametric combinations with $\mu = \mu_T$, while parametric combinations having $\mu = \mu_A$ were required to study the type II error α_2. Thus, for a given range of the necessarily prespecified values N_{min}, it was quite unlikely that any particular N_{min} would give error rates α_1', α_2' exactly equal to the specified values α_1, α_2. By proceeding in a trial-and-error manner it would have been possible to find N_{min} levels which very nearly gave the desired error rates, but practical time and economic limitations necessitated using censoring levels N_{min} which gave somewhat crude approximation to specified error rates. In general, the approximation was such that α_1' and α_2' were less than or equal to α_1 and α_2 for $\alpha_1 = \alpha_2 = 0.05$, while the reverse was true for $\alpha_1 = \alpha_2 = 0.01$. To give some indication of the order of this approximation we consider the results from two representative series of trials, each consisting of 200 trials, the parametric combinations being $(\mu,\mu_T,\mu_A,\rho) = (0,0,0.4,0.75)$ and $(\mu,\mu_T,\mu_A,\rho) = (0.4,0,0.4,0.75)$. For these parametric combinations, tests L_1 and L_2 were each censored at the tenth stage with the following results:

	L_1	L_2
α_1, α_2	α_1', α_2'	α_1', α_2'
0.05, 0.05	0.020, 0.040	0.020, 0.045
0.05, 0.01	0.020, 0.025	0.020, 0.025
0.01, 0.05	0.010, 0.040	0.015, 0.045

It may be remarked that discrepancies greater than those appearing in the second row above were in fact rare.

It should be noted that the results of the necessarily limited numerical investigation to be presented in the remainder of this section serve only as a guide to the behavior of the large-sample tests. However, the actual scope of the investigation complicated the task of reporting the results because so many tables would have been needed for an exhaustive presentation. To simplify the presentation, symmetry in results was taken advantage of by averaging (or pooling) these results and reporting them as one. For example, with boundary B_1 (for which α_1,α_2=0.05,0.05) the a.s.n. for each series of trials was approximately the same given H_T (i.e. when $\mu=\mu_T$) as when given H_A; so these numbers were averaged. The same was true for B_2 (where α_1,α_2 = 0.05,0.01) given H_T and B_3 (where α_1,α_2 = 0.01,0.05) given H_A and for B_3 given H_T and B_2 given H_A. By combining results in this manner, it is felt that the presentation is improved through the resultant simplification and is without serious distortion.

Table 1 gives censoring levels N_{min} for the large-sample test with covariance L_1 for the argument $|\mu_A - \mu_T|/\sigma$. The corresponding results for the large-sample test ignoring covariance L_2 are given in Table 2. The table headings state that each level of censoring is determined on the basis of 200 trials with $\mu = \mu_T$ and 200 trials with $\mu = \mu_A$. The meaning of this is that for each argument γ, 400 trials in all were conducted to find the level of censoring N_{min} which gave α_1', $\alpha_2' = \alpha_1$, α_2, 200 trials having $\mu = \mu_T$ to obtain α_1' and 200 trials having $\mu = \mu_A$ to obtain α_2'.

Table 1. Test L_1: censoring levels N_{min} for argument $\gamma_1 = |\mu_A - \mu_T|/\sigma$, each N_{min} determined from 200 trials with $\mu=\mu_T=0$ and 200 trials with $\mu=\mu_A$

μ_A	ρ	γ_1	N_{min}
0.20	0.75	0.30	25
0.40	0.75	0.60	14
0.72	0.60	0.90	10
0.96	0.60	1.20	6
1.20	0.60	1.50	5

Table 2. Test L_2: censoring levels N_{min} for argument $\gamma_2 = |\mu_A - \mu_T| / \sigma_y$, each N_{min} determined from 200 trials with $\mu=\mu_T=0$ and 200 trials with $\mu=\mu_A=\gamma_2$

γ_2	N_{min}
0.20	20
0.40	16
0.72	10
0.96	6
1.20	4

Inspection of Tables 1 and 2 shows that as γ increases, the necessary level of censoring decreases very rapidly. As was mentioned earlier, the minimum sample numbers for tests L_1 and L_2 are $n = 3$ and $n = 2$ respectively. Thus with γ large — say $\gamma \geq 1.2$ — censoring which delays a decision for only two or three stages would appear to be adequate. On the other hand, for γ small — say $\gamma \cong 0.6$ — it is necessary to delay decisions until around the fourteenth stage, and much longer as γ becomes yet smaller.

Table 3 presents the observed a.s.n., $\bar{n}$, for test L_1 against argument γ_1 where the $\bar{n}$ has been averaged for $B_1 | H_T$ and $B_1 | H_A$, for $B_2 | H_T$ and $B_3 | H_A$, and for $B_3 | H_T$ and $B_2 | H_T$

(where we use the notation $B_2 | H_T$, for example to denote B_2 given H_T). The analogous results are also given for the s.p.r.t. with covariance, S_1. Table 4 presents the corresponding results for the tests without covariance, L_2 and S_2. The tabular entries reflect the a.s.n.'s resulting from the censoring levels N_{min} given in Tables 1 and 2. Of course S_1 and S_2 were not censored.

Tables 5 and 6 are the $\hat{V}(n)$ analogues to Tables 3 and 4, where $\hat{V}(n)$ denotes the observed variance of the sample number.

Caution should be used in comparing the large-sample tests with the s.p.r.t.'s for in general, the s.p.r.t.'s seem to be somewhat conservative in that they appear to have greater power than for which they were designed. This means that the a.s.n.'s for the s.p.r.t.'s are greater than would result if the nominal error probabilities were more nearly realized. As was pointed out earlier, the large-sample tests designed to give error rates $\alpha_i = 0.05$ (i=1,2) are also somewhat conservative since in general $\alpha_i' \leq \alpha_i$. However the large-sample tests designed to give error rates $\alpha_i = 0.01$ in general require more censoring since $\alpha_i' \geq \alpha_i$. Also in

Table 3. Tests L_1 and S_1: a.s.n. for argument $\gamma_1 = |\mu_A - \mu_T|/\sigma$, each entry based on 400 trials

γ_1	L_1: $B_1 \mid (H_T$ or $H_A)$	L_1: $B_2 \mid H_T$ or $B_3 \mid H_A$	L_1: $B_3 \mid H_T$ or $B_2 \mid H_A$	S_1: $B_1 \mid (H_T$ or $H_A)$	S_1: $B_2 \mid H_T$ or $B_3 \mid H_A$	S_1: $B_3 \mid H_T$ or $B_2 \mid H_A$
0.30	61.0	90.8	65.6	61.0	95.0	66.0
0.60	19.6	25.4	20.6	17.2	25.6	19.2
0.90	11.8	13.6	12.0	8.6	12.5	8.8
1.20	7.1	8.1	7.2	5.0	6.9	5.4
1.50	5.6	6.2	5.6	3.4	4.8	3.6

Table 4. Tests L_2 and S_2: a.s.n. for argument $\gamma_2 = |\mu_A - \mu_T|/\sigma_y$, each entry based on 400 trials

γ_2	L_2: $B_1 \mid (H_T$ or $H_A)$	L_2: $B_2 \mid H_T$ or $B_3 \mid H_A$	L_2: $B_3 \mid H_T$ or $B_2 \mid H_A$	S_2: $B_1 \mid (H_T$ or $H_A)$	S_2: $B_2 \mid H_T$ or $B_3 \mid H_A$	S_2: $B_3 \mid H_T$ or $B_2 \mid H_A$
0.20	131.2	203.8	143.2	136.6	206.6	145.1
0.40	34.2	50.2	37.3	35.8	54.4	38.0
0.72	14.6	18.2	14.8	12.2	17.6	12.8
0.96	8.6	11.0	8.8	7.3	10.8	7.5
1.20	5.4	6.7	5.6	5.0	7.3	5.2

Table 5. Tests L_1 and S_1: $\hat{V}(n)$ for argument $\gamma_1 = |\mu_A - \mu_T|/\sigma$, each entry based on 400 trials

γ_1	L_1			S_1		
	$B_1 \mid (H_T$ or $H_A)$	$B_2 \mid H_T$ or $B_3 \mid H_A$	$B_3 \mid H_T$ or $B_2 \mid H_A$	$B_1 \mid (H_T$ or $H_A)$	$B_2 \mid H_T$ or $B_3 \mid H_A$	$B_3 \mid H_T$ or $B_2 \mid H_A$
0.30	1669	3242	2202	2036	3303	2638
0.60	97	217	123	134	233	210
0.90	15	32	17	34	54	35
1.20	6	14	8	10	14	14
1.50	2	4	2	5	7	5

Table 6. Tests L_2 and S_2: $\hat{V}(n)$ for argument $\gamma_2 = |\mu_A - \mu_T|/\sigma_y$, each entry based on 400 trials

γ_2	L_2			S_2		
	$B_1 \mid (H_T$ or $H_A)$	$B_2 \mid H_T$ or $B_3 \mid H_A$	$B_3 \mid H_T$ or $B_2 \mid H_A$	$B_1 \mid (H_T$ or $H_A)$	$B_2 \mid H_T$ or $B_3 \mid H_A$	$B_3 \mid H_T$ or $B_2 \mid H_A$
0.20	10136	17488	12910	11426	16250	13698
0.40	454	1076	646	583	1019	714
0.72	54	100	57	68	113	82
0.96	15	39	17	22	38	22
1.20	6	13	7	13	24	13

comparing $\hat{V}(n)$ for the large-sample tests and for the s.p.r.t.'s, it should be noted that censoring, in general, will tend to reduce $\hat{V}(n)$ since very small observations are eliminated.

With the above remarks in mind, it will be noted that on the basis of the Monte Carlo results, the large-sample tests have a.s.n.'s which are approximately equal to those of the corresponding s.p.r.t.'s and the $\hat{V}(n)$ are in general small for the large-sample tests.

The cautionary remarks made above are also pertinent when one attempts to make meaningful comparisons between tests which utilize concomitant information and those which do not, i.e. L_1 versus L_2 and S_1 versus S_2. Consider a given testing situation wherein all parameters and the base hypotheses H_T and H_A are fixed, and consider the tests L_1 and L_2 (with the proper censoring levels) and S_1 and S_2 which have all been designed to test the hypothesis H_T against H_A with error probabilities α_1, α_2. If all of the tests give exactly the same error rates α_1', $\alpha_2' = \alpha_1$, α_2, it is convenient to describe such tests as being "analogues." But in a given series of trials, these tests give only approximately the same error rates α_1', α_2' and are thus actually only "approximate analogues." The order of this approximation was indicated earlier. Further, it is noted that since

$$\sigma = \sigma_y(1-\rho^2)^{1/2} ,$$

a particular series of trials has associated arguments

$$\gamma_1 = \frac{|\mu_A - \mu_T|}{\sigma}$$

and

$$\gamma_2 = \frac{|\mu_A - \mu_T|}{\sigma_y} ,$$

and it follows that

$$\gamma_1 = \frac{\sigma_y}{\sigma} \gamma_2 \geq \gamma_2 \quad ,$$

with equality holding only when $\rho = 0$. Thus we may consider the pairs γ_1, γ_2 which correspond to the various series of trials; then meaningful comparisons may be made in Tables 3-6 by comparing results from tests which are "approximate analogues." The "approximate analogues" are indicated by the pairs γ_1, γ_2 in Table 7. For example, we see that the a.s.n.

Table 7. Pairs γ_1, γ_2 indicating "approximate analoges

γ_1, γ_2
0.30, 0.20
0.60, 0.40
0.90, 0.72
1.20, 0.96
1.50, 1.20

entries in Table 3 (for tests L_1 and S_1) for argument $\gamma_1 = 0.90$ and the a.s.n. entries in Table 4 (for tests L_2 and S_2) for argument $\gamma_2 = 0.72$ come from tests which are "approximate analogues."

By comparing only results from "approximate analagoues" (in an attempt to make meaningful comparisons) it is seen that the large-sample test utilizing concomitant information, L_1 gives appreciably smaller a.s.n. and $\hat{V}(n)$ than the large-sample test ignoring concomitant information, L_2. The same holds for the corresponding s.p.r.t.'s, S_1 and S_2.

In section A of this chapter, it was conjectured that approximately (see 4.5 and 4.6)

$$\frac{C_s}{N_s} = 1-\rho^2 ,$$

where C_s and N_s denote a.s.n.'s needed to detect $\mu_A - \mu_T$ as being significant in sequential tests with and without concomitant information respectively. Results of trials with "approximate analogues" are used to investigate this conjecture. Table 8 gives the appropriate a.s.n. ratios needed to investigate the conjecture for s.p.r.t.'s, S_1 and S_2. Table 9 gives the corresponding check for the large-sample tests L_1 and L_2. There are some multiple entries in these tables since more than one series of trials (with "approximate analogues") were available on some values of .

Table 8. Observed a.s.n. (S_1) to a.s.n. (S_2) ratios, each a.s.n. being based on 400 trials

ρ	Conjectured ratio $1-\rho^2$	γ_1, γ_2	$B_1 \mid (H_T$ or $H_A)$	$B_2 \mid H_T$ or $B_3 \mid H_A$	$B_3 \mid H_T$ or $B_2 \mid H_A$
0.60	0.64	0.90, 0.72	0.70	0.71	0.69
		1.20, 0.96	0.68	0.64	0.72
		1.50, 1.20	0.68	0.66	0.69
0.75	0.44	0.30, 0.20	0.45	0.46	0.45
		0.60, 0.40	0.48	0.47	0.50
0.90	0.19	0.92, 0.40	0.22	0.22	0.23

Table 9. Observed a.s.n. (L_1) to a.s.n. (L_2) ratios, each a.s.n. being based on 400 trials

ρ	Conjectured ratio $1-\rho^2$	γ_1, γ_2	$B_1 \mid (H_T$ or $H_A)$	$B_2 \mid H_T$ or $B_3 \mid H_A$	$B_3 \mid H_T$ or $B_2 \mid H_A$
0.60	0.64	0.90, 0.72	0.81	0.75	0.81
		1.20, 0.96	0.82	0.74	0.82
		1.50, 1.20	1.04	0.93	1.00
0.75	0.44	0.30, 0.20	0.46	0.46	0.46
		0.60, 0.40	0.57	0.51	0.55
0.90	0.19	0.92, 0.40	0.33	0.25	0.31

It is seen from Table 8, when S_1 and S_2 are compared, that the numerical results are in quite close agreement with the conjecture on relative a.s.n., but — in general — the gain from using concomitant information is slightly less than conjectured. The comparison based on large-sample tests did not give such close agreement with the conjecture, in general, as seen in Table 9. The agreement is not too bad on $\rho = 0.75$ but is particularly bad on $\rho = 0.60$. Part of the latter difficulty is due to trials in which γ_1 and γ_2 were both large so that censoring played a large role in determining the average sample numbers.

In summary, the results of the Monte Carlo suggest that the large-sample tests compare quite favorably with their respective s.p.r.t. counterparts on the bases of a.s.n. and $\hat{V}(n)$ (see Tables 3-6). Also, appreciable reduction in a.s.n. and $\hat{V}(n)$ may be realized by utilizing concomitant information (see Tables 3-9). There is, however, one major drawback associated with the large-sample tests; they must be censored, and — in the Monte Carlo trials described herein — we have used our knowledge of σ_y and σ to determine the appropriate level of censoring. In practice, of course, σ and σ_y will seldom be known. It was pointed out, however, that if $|\mu_A - \mu_T|$ is known to be "large" relative to the standard deviation very little censoring is required. In situations where the experimenter has some knowledge about the standard deviation, say from a pilot study, Tables 1 and 2 provide a rough guide to the amount of censoring required. In general, if the true standard deviation exceeds the value used to determine the level of censoring, γ will be overestimated and the error probabilities of the resulting test will exceed the predetermined values. When the true standard deviation is less than the value used to determine the level of censoring, the reverse will be true, i.e. an overly conservative test will result with an associated loss of efficiency.

In situations where the experimenter is in complete ignorance about the standard deviation, the optimum course of action is not clear. A possible approach, but one which is open to criticism, is the following: We note that even when γ is "large" (see Tables 1 and 2), decisions are delayed at least until around the fifth stage. Therefore, observations from the first five stages might be used to give a pilot estimate of the standard deviation. This approach is at best a precarious one, however, since the variance of the estimated standard deviation is large for n small. Indeed this condition contributes greatly to the need for censoring. Further work is needed before a strong recommendation can be made on this point.

C. Weight-Function Tests: Numerical Examples

In Chapter III the weight-function approach was used to develop sequential test procedures for utilizing concomitant information. Two hypothesis formulations were considered. Formulation 1 is

$$H_T: \mu=\mu_T; \; H_A: \mu=\mu_A$$

for one-sided alternatives, and

$$H_T: \mu=\mu_T; \; H_A: \mu=\pm\mu_A$$

for two-sided alternatives, where — for simplicity — we let $\pm\mu_A = \mu_T + d$ (d being a specified constant) when $\mu_T \neq 0$. In addition a modification of formulation 1,

$$H_T: \mu\leq\mu_T; \; H_A: \mu\geq\mu_A \quad (\mu_T<0<\mu_A) \; ,$$

was considered.

Formulation 2 is

$$H_T: \mu=\mu_T; \; H_A: \mu=\mu_T+\gamma\sigma$$

for one-sided alternatives and

$$H_T: \mu=\mu_T; \; H_A: |\mu-\mu_T|=\gamma\sigma \quad (\gamma \text{ being a specified constant } > 0)$$

for two-sided alternatives.

We shall designate the weight-function test procedures with the notation f(i,j,k), (i,j,k=1,2); here f(i,j,k) denotes the procedure for the i-th formulation with j-sided alternatives utilizing covariance denoted by k = 1 and without covariance denoted by k = 2. For example, f(1,2,1) denotes the test procedure for formulation 1 with two-sided alternatives when covariance is utilized, f(2,1,2) signifies the test procedure for formulation 2 with one-sided alternatives and ignoring covariance, etc. For the modification of formulation 1, we shall denote the associated test procedures by m(1,1) when covariance is used and by m(1,2) when covariance is ignored.

For reference, the n-th stage test statistics (P_n^*) corresponding to each test procedure are given below.

f(1,1,1) (see Equation 3.14):

$$P_n^* = \left[\frac{\Sigma x^2 \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{\Sigma x^2 \Sigma(y-\mu_A)^2 - [\Sigma(y-\mu_A)x]^2}\right]^{\frac{n-2}{2}} ; \qquad (4.13)$$

f(1,1,2):

$$P_n^* = \left[\frac{\Sigma(y-\mu_T)^2}{\Sigma(y-\mu_A)^2}\right]^{\frac{n-1}{2}} ; \qquad (4.14)$$

f(1,2,1) (see Equation 3.36):

$$P_n^* = \frac{1}{2}\left[\frac{\Sigma x^2\ \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{\Sigma x^2\ \Sigma(y-\mu_A)^2 - [\Sigma(y-\mu_A)x]^2}\right]^{\frac{n-2}{2}}$$

$$+ \frac{1}{2}\left[\frac{\Sigma x^2\ \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{\Sigma x^2\ \Sigma(y+\mu_A)^2 - [\Sigma(y+\mu_A)x]^2}\right]^{\frac{n-2}{2}} ; \qquad (4.15)$$

f(1,2,2):

$$P_n^* = \frac{1}{2}\left[\frac{\Sigma(y-\mu_T)^2}{\Sigma(y-\mu_A)^2}\right]^{\frac{n-1}{2}} + \frac{1}{2}\left[\frac{\Sigma(y-\mu_T)^2}{\Sigma(y+\mu_A)^2}\right]^{\frac{n-1}{2}} ; \qquad (4.16)$$

f(2,1,1) (see Equation 3.27):

$$P_n^* = e^c F\left(\frac{n-2}{2},\frac{1}{2};\frac{b^2}{4a}\right) - \frac{(n-3)!(\pi a)^{1/2} e^c}{2^{n-4}[(\frac{n-4}{2})!]^2 b}\left[F\left(\frac{n-1}{2},\frac{1}{2};\frac{b^2}{4a}\right) - F\left(\frac{n-3}{2},\frac{1}{2};\frac{b^2}{4a}\right)\right]$$

for n = 4,6,8,..., and

$$P_n^* = e^c F\left(\frac{n-2}{2},\frac{1}{2};\frac{b^2}{4a}\right) - \frac{2^{n-2}[(\frac{n-3}{2})!]^2 a^{1/2} e^c}{(n-3)!\,\pi^{1/2} b}\left[F\left(\frac{n-1}{2},\frac{1}{2};\frac{b^2}{4a}\right) - F\left(\frac{n-3}{2},\frac{1}{2};\frac{b^2}{4a}\right)\right] \qquad (4.17)$$

for n=5,7,9,..., where

$$a = \frac{\Sigma x^2\ \Sigma(y-\mu_T)^2 - [\Sigma(y-\mu_T)x]^2}{2\Sigma x^2} ,$$

$$b = -\gamma_1 \frac{[\Sigma x^2\ \Sigma(y-\mu_T) - \Sigma x(y-\mu_T)\ \Sigma x]}{\Sigma x^2} ,$$

and

$$c = -\gamma_1^2 \frac{[n\Sigma x^2 - (\Sigma x)^2]}{2\Sigma x^2} = -\frac{nS_{xx}}{2\Sigma x^2}\gamma_1^2 \ .$$

Also, $F(\frac{m}{2},\frac{1}{2};z)$ denotes the confluent hypergeometric function which is tabulated in (26).

f(2,1,2):

$$P_n^* = e^{c_1} F(\frac{n-1}{2},\frac{1}{2},\frac{b_1^2}{4a_1})$$

$$- \frac{2^{n-1}[(\frac{n-2}{2})!]^2 a_1^{1/2} e^{c_1}}{(n-2)!\pi^{1/2} b_1}[F(\frac{n}{2},\frac{1}{2},\frac{b_1^2}{4a_1})-F(\frac{n-2}{2},\frac{1}{2},\frac{b_1^2}{4a_1})]$$

for n = 4,6,8,..., and

$$P_n^* = e^{c_1} F(\frac{n-1}{2},\frac{1}{2},\frac{b_1^2}{4a_1})$$

$$- \frac{(n-2)!(\pi a_1)^{1/2} e^{c_1}}{2^{n-3}[(\frac{n-3}{2})!]^2 b_1}[F(\frac{n}{2},\frac{1}{2},\frac{b_1^2}{4a_1})-F(\frac{n-2}{2},\frac{1}{2},\frac{b_1^2}{4a_1})] \tag{4.19}$$

for n = 3,5,7,..., where

$$a_1 = \frac{1}{2}\Sigma(y-\mu_T)^2 \ ,$$

$$b_1 = -\gamma_2 \Sigma(y-\mu_T) \ , \tag{4.20}$$

and

$$c_1 = -\frac{n}{2}\gamma_2^2 ;$$

f(2,2,1) (see Equation 3.33)

$$P_n^* = e^c F\left(\frac{n-2}{2},\frac{1}{2};\frac{b^2}{4a}\right) , \qquad (4.21)$$

where a, b, and c are given by 4.18:

f(2,2,2) [Wald's sequential t-test as given by Arnold in (27, p. vi)]:

$$P_n^* = e^{c_1} F\left(\frac{n-1}{2},\frac{1}{2};\frac{b_1^2}{4a_1}\right) , \qquad (4.22)$$

where a_1, b_1, and c_1 are given by 4.20;

m(1,1) (see Equations 3.70-3.73):

$$P_4^* = -\frac{\mu_T}{\mu_A}\left[\frac{\pi-2\tan^{-1}[q^{-1}(2c_2\mu_A+b_2)]}{\pi+2\tan^{-1}[q^{-1}(2c_2\mu_T+b_2)]}\right] ,$$

$$P_5^* = -\frac{\mu_T}{\mu_A}\left[\frac{B(\mu_T)}{B(\mu_A)}\right]^{1/2}\left[\frac{2[c_2B(\mu_A)]^{1/2} - 2c_2\mu_A-b_2}{2[c_2B(\mu_T)]^{1/2} + 2c_2\mu_T+b_2}\right] ,$$

$$P_n^* = -\frac{\mu_T}{\mu_A}\left[\frac{\pi q^{-1/2} k^{\frac{n-4}{2}} [(n-4)!] - 2^{n-4}[(\frac{n-4}{2})!]^2 I_n(\mu_A)}{\pi q^{-1/2} k^{\frac{n-4}{2}} [(n-4)!] + 2^{n-4}[(\frac{n-4}{2})!]^2 I_n(\mu_T)}\right]$$

for $n = 6,8,10,\ldots,$ and

$$P_n^* = -\frac{\mu_T}{\mu_A}\left[\frac{c_2^{1/2} q^{-1} k^{\frac{n-5}{2}} 2^{n-3}[(\frac{n-5}{2})!]^2 - (n-4)[(n-5)!] I_n(\mu_A)}{c_2^{1/2} q^{-1} k^{\frac{n-5}{2}} 2^{n-3}[(\frac{n-5}{2})!]^2 + (n-4)[(n-5)!] I_n(\mu_T)}\right]$$

for $n = 7,9,11,\ldots,$ where

$$a_2 = \sum^n x^2 \sum^n y^2 - (\sum^n xy)^2 ,$$

$$b_2 = 2(\sum^n x \sum^n xy - \sum^n x^2 \sum^n y) ,$$

$$c_2 = n\sum^n x^2 - (\sum^n x)^2 ,$$

$$q = 4a_2c_2 - b2 ,$$

$$k = 4c_2\, q^{-1} ,$$

$$B(\mu) = \sum^n x^2 \sum^n (y-\mu)^2 - [\sum^n x(y-\mu)]^2$$

$$= c_2\mu^2 + b_2\mu + a_2 ,$$

$$I_4(\mu) = 2q^{-1/2} \tan^{-1}[q^{-1/2}(2c_2\mu + b_2)]$$

$$I_5(\mu) = 2(2c_2\mu + b_2)\, q^{-1}[B(\mu)]^{-1/2} ,$$

and, for n = 6,7,8,...,

$$I_n(\mu) = \frac{2(2c_2\mu+b_2)}{(n-4)q[B(\mu)]^{\frac{n-4}{2}}} + \frac{k(n-5)}{(n-4)} I'_{n-2}(\mu) . \qquad (4.24)$$

In the recursion formula for $I_n(\mu)$ in 4.24, the prime (') denotes that the quantities c_2, b_2, q,k, and $B(\mu)$ are based on sample size n. Also in 4.24, the quantity $\tan^{-1}[q^{-1/2}(2c_2\mu + b_2)]$ is the "principal value," i.e. the solution lying between $-\pi/2$ and $\pi/2$ as was discussed in Section E, Chapter III.

m(1,2):

$$P_3^* = -\frac{\mu_T}{\mu_A}\left[\frac{\pi-2\tan^{-1}[p^{-1/2}2(3\mu_A-\Sigma y)]}{\pi+2\tan^{-1}[p^{-1/2}2(3\mu_T-\Sigma y)]}\right] ,$$

$$P_4^* = -\frac{\mu_T}{\mu_A}\left[\frac{D(\mu_T)}{D(\mu_A)}\right]^{1/2}\left[\frac{2[D(\mu_A)]^{1/2} - 4\mu_A+\Sigma y}{2[D(\mu_T)]^{1/2} + 4\mu_T-\Sigma y}\right] ,$$

$$P_n^* = -\frac{\mu_T}{\mu_A}\left[\frac{\pi p^{-1/2} r^{\frac{n-3}{2}}[(n-3)!] - 2^{n-3}[(\frac{n-3}{2})!]^2 J_n(\mu_A)}{\pi p^{-1/2} r^{\frac{n-3}{2}}[(n-3)!] + 2^{n-3}[(\frac{n-3}{2})!]^2 J_n(\mu_T)}\right]$$

for $n = 5,7,9,\ldots,$ and

$$P_n^* = -\frac{\mu_T}{\mu_A}\left[\frac{n^{1/2} p^{-1} r^{\frac{n-4}{2}} 2^{n-2}[(\frac{n-4}{2})!]^2 - (n-3)[(n-4)!]J_n(\mu_A)}{n^{1/2} p^{-1} r^{\frac{n-4}{2}} 2^{n-2}[(\frac{n-4}{2})!]^2 + (n-3)[(n-4)!]J_n(\mu_T)}\right]$$

for $n = 6,8,10,\ldots,$ where

$$p = 4[n\Sigma y^2 - (\Sigma y)^2] = 4n\, S_{yy} ,$$

$$r = S_{yy}^{-1}$$

$$D(\mu) = \overset{n}{\Sigma}(y-\mu)^2 ,$$

$$J_3(\mu) = 2p^{-1/2} \tan^{-1}[p^{-1/2}(2c_3\mu + b_3)] ,$$

$$c_3 = n ,$$

$$b_3 = -2\overset{n}{\Sigma} y ,$$

$$J_4(\mu) = \frac{c(2c_3\mu+b_3)}{p[D(\mu)]^{1/2}}$$

and for n=5,6,7,...,

$$J_n(\mu) = \frac{2(2c_3\mu+b_3)}{(n-3)q[D(\mu)]^{\frac{n-3}{2}}} + \frac{(n-4)}{(n-3)S_{yy}} J'_{n-2}(\mu) \quad . \quad (4.26)$$

To ensure correct interpretation of the recursion formulae in 4.24 and 4.26 needed for tests m(1,1) and m(1,2), we consider an example. It is sufficient to illustrate the proper procedure for m(1,1) since the correct procedure for m(1,2) follows in the analogous way.

Assume that we have taken n = 8 observations x,y so that we wish to calculate P_8^*, and hence require $I_8(\mu_A)$ and $I_8(\mu_T)$. The correct use of the recursion formula in 4.24 is to calculate

$$I_8(\mu) = \frac{2(2c_2\mu+b_2)}{(8-4)q[B(\mu)]^{\frac{8-4}{2}}} + \frac{k(8-5)}{(8-4)} I'_6(\mu) \quad ,$$

where c_2, b_2, q, k, and $B(\mu)$ are all based on n = 8 observations. To compute $I'_6(\mu)$ we again use the recursion formula and obtain

$$I_6'(\mu) = \frac{2(2c_2\mu+b_2)}{(6-2)q[B(\mu)]^{\frac{6-4}{2}}} + \frac{k(6-5)}{(6-4)} I_4'(\mu) ,$$

where

$$I_4'(\mu) = 2q^{-1/2} \tan^{-1}[q^{-1/2}(2c_2\mu+b_2)] ,$$

and the quantities c_2, b_2, q, k, and $B(\mu)$ appearing in the expressions for $I_6'(\mu)$ and $I_4'(\mu)$ are again based on n = 8 observations (and not n = 6 and n = 4 respectively). In the corresponding way the recursion formula in 4.26 is used to obtain $J_8(\mu)$ in calculating the test statistic for test m(1,2).

Monte Carlo trials have not yet been conducted on the weight-function tests. These tests have been used and the test statistics are exhibited for a simple example which follows.

Table 10 gives 8 bivariate observations (x,y) generated from a bivariate normal population having $\mu_x = \mu_y = 0$, $\sigma_x = \sigma_y = 1$, and $\rho = 0.60$. The observations were obtained by the method described in Section B of this chapter.

The weight-function test procedures were used to analyze the data in Table 10 where in μ_T and μ_A were taken to be 0 and 1.2 respectively in formulation 1, so that γ_1, γ_2 become 1.5, 1.2 for formulation 2.

Table 10. Bivariate observations from the bivariate normal population having $\mu_x=\mu_y=0$, $\sigma_x=\sigma_y=1$, and $\rho=0.60$

Stage number n	x_n	y_n
1	-0.9423	-0.2365
2	0.9119	0.0942
3	0.4100	0.1209
4	1.1903	0.9693
5	-0.0831	-1.5913
6	0.3202	-0.4569
7	1.2796	0.7861
8	-0.3491	0.1692

It is recalled that in the modification of formulation 1, we required $\mu_T < 0 < \mu_A$ so that the weight functions would be properly defined (see Section E, Chapter III). As was indicated earlier, the test of $\mu \leq 0$ versus $\mu \geq 1.2$ on y may be obtained equivalently by testing, for example, $\mu \leq -1$ versus $\mu \geq 0.2$ on $y' = y-1$. Hence, test procedures m(1,1) and m(1,2) were performed on the data of Table 10 with y being replaced by y-1 and with μ_T, μ_A taken to be -1, 0.2.

The test results are reported in Table 11 which gives the value of each test statistic at each stage until termination. In conducting the tests, α_1, α_2 were taken to be 0.05, 0.05 so that the boundaries were

$$\frac{\alpha_2}{1-\alpha_1} = 0.053, \quad \frac{1-\alpha_2}{\alpha_1} = 19 .$$

Then, each test procedure was as follows:

At the n-th stage, the test statistic P_n^* was computed. If $P_n^* \leq 0.053$, the procedure was terminated and H_T was accepted. If $P_n^* \geq 19$, the procedure was terminated and H_A was accepted. If $0.053 < P_n^* < 19$, no decision was made and we continued to the next stage.

Inspection of Table 11 shows that all of the test procedures based on formulation 1 accepted H_T (correctly) at the third stage (it is noted that $\mu_A - \mu_T$ is relatively large). The computations of these test statistics were continued through further stages out of academic interest, and to compare their behavior with other test statistics which could not be calculated until subsequent stages. For it was seen in Chapter III that some of the test procedures were not defined for n small. For example, from Equation 4.17 we see that P_n^* for f(2,1,1) is not defined for n = 3 since $F(0,\frac{1}{2};\frac{b^2}{4a})$ is not defined. Also, for n small, some entries in Table 10 are blank because the hypergeometric function $F(\frac{m}{2},\frac{1}{2};z)$ is not tabulated in (26) for m < 3. Hence, from 4.17 and 4.19 we see that P_4^* and P_5^* for f(2,1,1) could not be calculated from (26), nor could P_3^* and P_4^* for

Table 11. Test statistics P_n^* calculated from the data in Table 10

Test procedure	Stage number n					
	3	4	5	6	7	8
f(1,1,1)	0.051	0.073	0.105	0.043		
f(1,1,2)	0.018	0.108	0.083	0.031		
f(1,2,1)	0.052	0.066	0.220	0.164	0.106	0.018
f(1,2,2)	0.018	0.073	0.116	0.062	0.028	
m(1,1)	---	0.571	0.330	0.091	0.036	
m(1,2)	0.407	0.914	0.346	0.087	0.082	0.032
f(2,1,1)	---	---	---	0.001		
f(2,1,2)	---	---	0.013			
f(2,2,1)	---	---	0.043			
f(2,2,2)	---	0.239	0.038			

f(2,1,2). Similarly, Equations 4.21 and 4.22 show that P_3^* and P_4^* for f(2,2,1) and P_3^* for f(2,2,2) could not be calculated using the tables in (26). From Table 11 we see that test procedures f(1,1,1) and f(1,1,2), having accepted H_T on stage 3 do not again accept H_T until stage 6. Test procedure f(1,2,1), after accepting H_T on stage 3, does not accept H_T again until stage 8, and test f(1,2,2) does not re-accept H_T until stage 7.

A point made in passing is that for test procedure f(2,2,2) it is possible to calculate a lower bound on the sample number required to accept H_T. H_T is accepted when

$$P_n^* \le \frac{\alpha_2}{1-\alpha_1}$$

From 4.20 and 4.22 we see that this is the condition that

$$\log P_n^* = c_1 + \log F\left(\frac{n-1}{2},\frac{1}{2};\frac{b_1^2}{4a_1}\right)$$

$$= -\frac{n}{2}\gamma_2^2 + \log F\left(\frac{n-1}{2},\frac{1}{2};\frac{b_1^2}{4a_1}\right)$$

$$\le \log \frac{\alpha_2}{1-\alpha_1}$$

Since $F\left(\frac{n-1}{2},\frac{1}{2};\frac{b_1^2}{4a_1}\right) \ge 1$ (with equality holding only for $\frac{b_1^2}{4a_1} = 0$) it follows that H_T cannot be accepted until

$$n \ge \frac{-2}{\gamma_2^2} \log \frac{\alpha_2}{1-\alpha_1} \quad .$$

For the case in point, f(2,2,2) cannot accept H_T until $n \ge 4.09$, i.e. until stage 5.

The corresponding bound for test f(2,2,1) is

$$n \ge -\frac{2\Sigma x^2}{S_{xx}\gamma_1^2} \log \frac{\alpha_2}{1-\alpha_1} ,$$

but this is less useful since it depends on sample observations (x).

We now discuss an apparent difficulty associated with the test procedures for formulation 1. Consider test procedure f(1,1,2) with P_n^* given by 4.14. Writing $y-\mu_A+\mu_A-\mu_T$ for $y-\mu_T$ we obtain

$$\frac{\Sigma(y-\mu_T)^2}{\Sigma(y-\mu_A)^2} = 1 + \frac{2(\mu_A-\mu_T)n(\bar{y}-\mu_A)+n(\mu_A-\mu_T)^2}{\Sigma(y-\mu_A)^2} \tag{4.27}$$

If we write $y-\bar{y}+\bar{y}-\mu_A$ for $y-\mu_A$, 4.27 becomes

$$\frac{\Sigma(y-\mu_T)^2}{\Sigma(y-\mu_A)^2} = 1 + \frac{2n(\mu_A-\mu_T)(\bar{y}-\frac{\mu_A+\mu_T}{2})}{\Sigma(y-\bar{y})^2-2(\bar{y}-\mu_A)\Sigma(y-\bar{y})+n(\bar{y}-\mu_A)^2}$$

$$= 1 + \frac{2(\mu_A-\mu_T)(\bar{y}-\frac{\mu_A+\mu_T}{2})}{S^2+(\bar{y}-\mu_A)^2}, \tag{4.28}$$

where

$$S^2 = \frac{\Sigma(y-\bar{y})^2}{n} . \tag{4.29}$$

Hence, letting

$$f(\bar{y}) = \frac{2(\mu_A-\mu_T)(\bar{y}-\frac{\mu_A+\mu_T}{2})}{S^2+(\bar{y}-\mu_A)^2}, \tag{4.30}$$

4.14 becomes

$$P_n^* = \left[\frac{\Sigma(y-\mu_T)^2}{\Sigma(y-\mu_A)^2} \right]^{\frac{n-1}{2}} = [1+f(\bar{y})]^{\frac{n-1}{2}} \quad . \tag{4.31}$$

If then S^2 is fixed, we have from 4.30,

$$\lim_{\bar{y} \to \infty} f(\bar{y}) = \lim_{\bar{y} \to -\infty} f(\bar{y}) = 0 \quad ,$$

so that upon holding n and S^2 fixed, 4.31 gives

$$\lim_{\bar{y} \to \infty} P_n^* = \lim_{\bar{y} \to -\infty} P_n^* = 1 \ . \tag{4.32}$$

The practical significance of 4.32 is that P_n^* tends toward unity when the true mean μ differs greatly from both μ_T and μ_A; therefore the test procedure would tend to continue much longer before accepting either H_T or H_A. A simple example illustrates this behavior. Suppose we are testing $H_T:\mu=0$ against $H_A:\mu=2$ with $\alpha_1=\alpha_2=0.01$. If we observe six successive observations of y=3, we have

$$P_6^* = 9^{5/2} = 243 > 99$$

and the test procedure terminates with acceptance of H_A. However, if we observe successive observations of y=10, we require 23 stages to accept H_A which is surprising and apparently inconsistent with the earlier acceptance in the previous case.

It has not been substantiated analytically whether this behavior obtains in the formulation 1 test procedures utilizing concomitant information, but their parallel development might indicate the affirmative. Close examination of the basis on which these test procedures were developed is helpful in considering this point. For illustrative purposes consider test f(1,1,2). It is seen from Sections A and B of Chapter III that the parameter space $\Delta = (\mu,\beta,\sigma)$ was divided into mutually exclusive regions, the region of preference for H_T, $\Delta_a = (\mu_T,\beta,\sigma)$, the region of preference for H_A, $\Delta_r = (\mu_A,\beta,\sigma)$, and the region of indifference which we may denote as $\Delta_I = (\mu_T<\mu<\mu_A,\beta,\sigma)$, where for simplicity we are assuming $\mu_A > \mu_T$. In this framework, we have made no specific provision for the cases when $\mu > \mu_A$ and $\mu < \mu_T$. Hence it is not too surprising to find that when $\bar{y}$ tends to differ appreciably from both μ_T and μ_A (that is either $\bar{y} \ll \mu_T < \mu_A$ or $\mu_T < \mu_A \ll \bar{y}$) the sample number required to reach a decision tends to increase. For indeed the sample evidence in such a situation tends to show a preference for neither $H_T: \mu=\mu_T$ nor $H_A: \mu=\mu_A$. As a result, efficient usage of the formulation 1 test procedures requires critical assessment of the base hypotheses H_T and H_A. If neither μ_T nor μ_A is close to the true mean μ, an unduly large sample might be required before a decision can be made.

To avoid the above discussed difficulty, we considered the modification of formulation 1,

$$H_T: \mu \leq \mu_T;\ H_A: \mu \geq \mu_A \quad (\mu_T < 0 < \mu_A).$$

Thus, for test m(1,1), the division of the parameter space $\Delta = (\mu,\beta,\sigma)$ is given by $\Delta_a = (\mu \leq \mu_T, \beta, \sigma)$, $\Delta_r = (\mu \geq \mu_A, \beta, \sigma)$, and $\Delta_I = (\mu_T < \mu < \mu_A, \beta, \sigma)$ so that the regions are mutually exclusive and exhaustive.

It is not anticipated that this problem carries over into the formulation 2 procedures, or at least into these procedures developed for two-sided alternatives. The test statistic for procedure f(2,2,2) is in fact that given by Arnold in (27, p. vi) for Wald's sequential t-test, this test being optimum in the sense discussed in Section A, Chapter III. To develop this test, Wald (28, p. 83) divided the parameter space into three mutually exclusive and exhaustive regions — $\Delta_a = (\mu_T, \sigma_y)$ $\Delta_r = (\mu$ such that $|\mu - \mu_T| \geq \gamma_2\sigma_y, \sigma_y)$, and $\Delta_I = (\mu$ such that $|\mu - \mu_T| < \gamma_2\sigma_y, \sigma_y)$. The close parallel in the developments of tests f(2,2,2) and f(2,2,1) and the resulting similarities in the form of their test statistics P_n^* (see Equations 4.21 and 4.22) might suggest that f(2,2,1) is free of this problem. But this can be no more than a conjecture due to the lack of substantiating theory at present and in absence of Monte Carlo experience. It is hoped to subject this point to further examination.

V. SUMMARY AND CONCLUDING REMARKS

A. Summary

The problem of utilizing measured concomitant information in sequential trials for comparing two treatments has been studied. Successive subjects are paired, each member of each pair being randomly allocated to one of the two treatments. The analysis is based upon the treatment 1 minus treatment 2 response differences $y=y_1-y_2$ and the corresponding concomitant observation differences $x=x_1-x_2$, where x and y are assumed to follow the bivariate normal distribution with parameters $\mu_y=\mu$, $\mu_x=0$, σ_x^2, σ_y^2, and correlation coefficient ρ. Alternatively, it follows that we may consider the problem as being that of sequentially testing hypotheses about the mean of a normal variable y when a normally distributed variable x, having mean zero and being correlated with y, may be observed.

Sequential procedures have been developed for two basic formulations of the test hypotheses. Formulation 1 is $H_T:\mu=\mu_T$ versus $H_A:\mu=\mu_A$ for one-sided alternatives and $H_T:\mu=\mu_T$ versus $H_A:\mu\pm\mu_A$ for two-sided alternatives, where — for simplicity — we let $\pm\ \mu_A$ denote $\mu_T\pm d$, d being a specified constant. In addition a modification of formulation 1, $H_T:\mu\leq\mu_T$ versus $H_A:\mu\geq\mu_A$ ($\mu_T<\mu_A$) is considered. Formulation 2 is $H_T:\mu=\mu_T$ versus $H_A:\mu=\mu_T+\gamma\sigma$ for one-sided alternatives and

$H_T: \mu = \mu_T$ versus $H_A: |\mu - \mu_T| = \gamma\sigma$ for two-sided alternatives, γ being a specified constant.

The hypotheses under consideration are composite since nuisance parameters are present. Hence, Wald's s.p.r.t. (sequential probability ratio test), which provides a satisfactory solution for testing simple hypotheses, is not directly applicable. Two approaches to the problem are proposed which do, however, make use of the basic Wald framework.

In Chapter II a large-sample test procedure is developed for formulation 1. It is shown that the Wald-type test in which nuisance parameters are replaced by maximum likelihood estimates is asymptotically equivalent to the s.p.r.t. (in which these parameters are specified). Without modification, however, this large-sample test appears to decision prematurely in small samples resulting in excessive error realizations on the base hypotheses. Results from an empirical investigation, which are reported in Chapter IV, suggest that the large-sample test may be "censored" (by allowing termination only after a specified stage) to obtain satisfactory error rates. However the appropriate degree of censoring appears to be dependent in a not yet elucidated way upon μ_A, μ_T, and σ. The results of Monte Carlo trials in which censoring, average sample number, and variance of sample number are studied for specified values of $|\mu_A - \mu_T|/\sigma$ are given for the large-sample procedure as for the corresponding large-sample procedure which ignores concomitant information. Monte Carlo results on a conjecture concerning the reduction of sample size due to the utilization of concomitant information are also reported.

The Monte Carlo results showed that the censored large-sample test procedures in general had average sample numbers of the same order as those of their corresponding s.p.r.t.'s while the observed variance of the sample number for each large-sample procedure was smaller than that of the corresponding s.p.r.t. Further, it was seen that approciable gains resulted from utilizing concomitant information, although the gains were slightly less in general than had been conjectured.

In Chapter III sequential procedures for both basic formulations of the hypotheses and for the modification of formulation 1 were developed by using Wald's weight-function approach to the problem of composite hypothesis testing. In essence, this approach involves specifying parametric weight-functions so that in turn a form of a prior (parametric) distribution is determined. Then the s.p.r.t. based upon this prior distribution is such that the "weighted average" of the error probabilities for parametric points lying in the acceptance-rejection regions correspond to specified theoretical values. Although Wald (28, pp. 204-207) showed that "optimum" weight functions exist for the sequential t-test, that is weight functions which assure that nominal theoretical error probabilities are not exceeded for any parametric points in the acceptance and rejection regions, no general existence theorems are available. It has not been determined whether the weight-functions used in Chapter III are optimum.

Although the associated computations are somewhat laborious by desk calculator, the test statistic for each weight-function procedure has been presented either in an explicit form or in a form which makes use of existing tables.

B. Concluding Remarks

We now examine briefly the correlation structure in general and as it pertains to our model assumptions given in Chapter I. In situations where treatment responses and concomitant observations are highly correlated, we propose to advantageously use this correlation when comparing two treatments. The technique involves pairing successive observations and analyzing treatment 1 minus treatment 2 response differences $y=y_1-y_2$ and the corresponding concomitant observation differences $x=x_1-x_2$ where y,x are assumed to be bivariate normally distributed with parameters $\mu_y=\mu$, $\mu_x=0$, σ_x^2, σ_y^2, and correlation coefficient ρ.

We define the vector $\underline{z}$ to be

$$\underline{z}' = (x_1\ x_2\ y_1\ y_2) \tag{5.1}$$

and consider the case where the covariance structure is given as

$$\mathrm{Cov}\,(\underline{z},\underline{z}') = \begin{bmatrix} \sigma_x'^2 & \rho_x\sigma_x'^2 & \rho_1\sigma_x'\sigma_y' & \rho_2\sigma_x'\sigma_y' \\ & \sigma_x'^2 & \rho_2\sigma_x'\sigma_y' & \rho_1\sigma_x'\sigma_y' \\ & & \sigma_y'^2 & \rho_y\sigma_y'^2 \\ & & & \sigma_y'^2 \end{bmatrix} ,$$

i.e.

$$\sigma_x' = \sigma_{x_i} \quad (i=1,2) ,$$

$$\sigma_y' = \sigma_{y_i} \quad (i=1,2) ,$$

$$\rho_1 = \rho_{x_i,y_j} \quad \text{for } i=j \quad (i,j=1,2) ,$$

$$\rho_2 = \rho_{x_i,y_j} \quad \text{for } i\neq j \quad (i,j=1,2) ,$$

$$\rho_x = \rho_{x_1,x_2} ,$$

and

$$\rho_y = \rho_{y_1,y_2} . \tag{5.2}$$

It is noted that in practice the pairing of successive observations might induce positive correlation between x_1 and x_2 (so that $\rho_x>0$) and between y_1 and y_2 (so that $\rho_y>0$).

Since $y=y_1-y_2$ and $x=x_1-x_2$ we have

$$\rho\sigma_x\sigma_y = \text{Cov }(x,y)=E[(x_1-Ex_1)-(x_2-Ex_2)][(y_1-Ey_1)-(y_2-Ey_2)]$$

$$= 2\,\rho_1\sigma_x'\sigma_y' - 2\,\rho_2\sigma_x'\sigma_y' = 2\sigma_x'\sigma_y'(\rho_1-\rho_2) \tag{5.3}$$

Then, since

$$\sigma_x^2 = \sigma_{x_1-x_2}^2 = 2\sigma_x'^2(1-\rho_x)$$

and

$$\sigma_y^2 = \sigma_{y_1-y_2}^2 = 2\sigma_y'^2(1-\rho_y) \ ,$$

Equation 5.3 may be written as

$$\rho = (\rho_1-\rho_2)[(1-\rho_x)(1-\rho_y)]^{-1/2} \ . \tag{5.4}$$

In Chapter I, observations were assumed to be independent and so with x_1 and x_2 independent and with y_1 and y_2 independent, it follows that $\rho_x=\rho_y=\rho_2=0$ giving (from 5.4)

$$\rho = \rho_1 \ . \tag{5.5}$$

Thus the correlation for the differences x, y is precisely that correlation exhibited between x_1 and y_1 (and between x_2 and y_2) of which we wish to take advantage.

If, however, subjects do not enter the study in a random manner, the other correlations may be unequal to zero. Then ρ is the function of ρ_1, ρ_2, ρ_x, and ρ_y given in 5.4 and we see that it may exceed or be less than ρ_1.

Within the covariance structure given in 5.2, we have the cases

$$\rho_1, \rho_2 < 0, \quad \rho_x, \rho_y > 0 , \tag{5.8}$$

$$\rho_1, \rho_2 > 0, \quad \rho_x, \rho_y > 0 , \tag{5.9}$$

$$\rho_1 > 0, \ \rho_2 < 0, \ \rho_x, \rho_y < 0 , \tag{5.10}$$

and

$$\rho_1 < 0, \ \rho_2 > 0, \ \rho_x, \rho_y < 0 , \tag{5.11}$$

amongst which the cases of immediate practical interest are

$$\rho_1 > \rho_2 > 0, \ \rho_x, \rho_y > 0 \tag{5.12}$$

and

$$\rho_1 < \rho_2 < 0, \ \rho_x, \rho_y > 0 . \tag{5.13}$$

From 5.4 we see that in the cases 5.12 and 5.13 the tendency for $|\rho|$ to be "small" when ρ_1 and ρ_2 have the same sign tends to be offset by $[(1-\rho_x)(1-\rho_y)]^{-1/2}$ being "large" since ρ_x and ρ_y are positive. It would appear that a thorough study of the effect of correlated observations on the magnitude of $|\rho|$ would require more knowledge of the exact or relative magnitudes of the absolute values of ρ_1, ρ_2, ρ_x, and ρ_y. Other than that, it is possible to put somewhat loose bounds on $|\rho|$ by considering, for example, cases 5.12 and 5.13 wherein $\rho_x = \rho_y = \rho'$, and $|\rho_1| \geq |\rho'| \geq |\rho_2|$, say. We then obtain the bounds

$$\max\left[0, \frac{|\rho_1| - |\rho_2|}{1 - |\rho_2|}\right] \leq \frac{|\rho_1| - |\rho_2|}{1 - \rho'} = |\rho| \leq \min\left[1, \frac{|\rho_1| - |\rho_2|}{1 - |\rho_1|}\right].$$

There are several other important areas which require further research and it is of interest to discuss some of these in the light of the progress made. First, knowledge of the properties of the sequential procedures developed herein is very limited. For that matter, very little is known about the properties of other sequential tests for composite hypotheses in general. In fact, Johnson (16), in his extensive survey of sequential analysis, has pointed out that in most situations we have not even approximate formulae for the average sample number and operating characteristic functions. Some authors have reported Monte Carlo experience in lieu of theoretical derivations and useful information can be obtained in this way. The results of Monte Carlo trials with the large-sample procedures of Chapter II have been reported and there is a need for at least a comparable numerical investigation of the weight-function tests of Chapter III. To accomplish this, however, a study of simplifying transformations and/or the construction of special tables might be required due to the complexity of the associated test statistics.

Secondly, other approaches to the problem could perhaps be profitably studied. In particular, Jackson and Bradley (15) report on another approach to testing composite hypotheses which has found fairly general usage. It is the method of frequency functions proposed by Goldberg as described, for example, by Nandi (19). This method makes use of sufficient statistics having distributions not dependent on nuisance parameters. Following this approach, Rushton (20, 21) and Arnold in (27) independently developed a sequential t-test. In the frequency function method, successive values of the test statistic are taken to be the observations; hence, the observations are dependent and much of Wald's work does not apply. Barnard (5, 6) and Cox (11) establish conditions under which the frequency function of a test statistic may be used in a s.p.r.t. so that the expected error probabilities remain approximately at the predetermined levels. However, Jackson and Bradley (15) have found the verification of these conditions to be nontrivial and it has usually been omitted by authors using the method. A study of this method in conjunction with the current problem would be of interest.

A third important area for future study is that of truncation or closure. The sequential procedures discussed in this thesis are of an "open" design in that there is no finite upper bound on the sample number required to reach a decision. The possibility of exceptionally large sample sizes is somewhat undesirable in general, and in some cases the uncertainty of the point of termination may vitiate the use of sequential techniques. For example, knowledge of at least the maximum sample size may be needed for administrative reasons. For this reason some authors have constructed "closed" (or "truncated") designs in which termination prior to or at the N-th stage, say, is assured. Armitage (2, p. 30) reports that the first author to suggest specially constructed closed designs (rather than arbitrarily truncated open designs) was Bross (8). A more general class of closed designs was introduced by Armitage (1), and using some of the theory developed by Armitage in that paper, Schneiderman and Armitage (22, 23) described closed designs for univariate tests concerning the mean of a normal population with known and unknown variance and presented results from Monte Carlo trials. This work should provide a useful starting point for an investigation of truncated weight-function tests utilizing concomitant information.

There are yet other areas which merit study. One, for example, is the extension of results to the case of several concomitant variables. Another is the consideration of more than two treatments as in sequential analysis of covariance and/or sequential multiple comparisons which utilize available concomitant information.

VI. LITERATURE CITED

1. Armitage, P. Restricted sequential procedures. Biometrika 44: 9-26. 1957.

2. Armitage, P. Sequential medical trials. Oxford, England, Blackwell Scientific Publications. c1960.

3. Armitage, P. Sequential tests in prophylactic and therapeutic trials. Quarterly Journal of Medicine 23: 255-274. 1954.

4. Armitage, P. Some sequential tests of student's hypothesis. Royal Statistical Society Journal Supplement 9: 250-263. 1947.

5. Barnard, G. A. The frequency justification of certain sequential tests. Biometrika 39: 144-150. 1952.

6. Barnard, G. A. The frequency justification of sequential tests: addendum. Biometrika 40: 468-469. 1953.

7. Bateman, Harry. Tables of integral transforms. Vol. 1. New York, N. Y., McGraw-Hill Book Company, Inc. 1954.

8. Bross, I. Sequential medical plans. Biometrics 8: 188-205. 1952.

9. Cochran, William G. Analysis of covariance: its nature and uses. Biometrics 13: 261-281. 1957.

10. Cox, D. R. Large sample sequential tests for composite hypotheses. Sankhya 25: 5-12. 1963.

11. Cox, D. R. Sequential tests for composite hypotheses. Cambridge Philosophical Society Proceedings 48: 290-299. 1952.

12. Cramér, Harald. Mathematical methods of statistics. Princeton, N. J., Princeton University Press. 1961.

13. Davies, Owen L. The design and analysis of industrial experiments. New York, N. Y., Hafner Publishing Company. 1954.

14. International Business Machine Corp. Reference manual; random number generation and testing: Manual No. C20-8011. White Plains, N. Y., author. c1959.

15. Jackson, J. Edward and Bradley, Ralph A. Sequential X^2 and T^2 tests. Annals of Mathematical Statistics 32: 1063-1077. 1961.

16. Johnson, N. L. Sequential analysis: a survey. Royal Statistical Society Journal Series A, 24: 372-411. 1961.

17. Kendall, Maurice G. and Stuart, Alan. The advanced theory of statistics. 2nd ed. Vol. 1. New York, N. Y., Hafner Publishing Company. 1963.

18. Mann, H. B. and Wald, Abraham. On stochastic limit and order relationships. Annals of Mathematical Statistics 14: 217-226. 1943.

19. Nandi, H. K. Use of well-known statistics in sequential analysis. Sankhyā 8: 339-344. 1948.

20. Rushton, S. On a sequential t-test. Biometrika 37: 326-333. 1950.

21. Rushton, S. On a two-sided sequential t-test. Biometrika 39: 302-308. 1952.

22. Schneiderman, M. A. and Armitage, P. Closed sequential t-tests. Biometrika 49: 359-366. 1962.

23. Schneiderman, M. A. and Armitage, P. A family of closed sequential procedures. Biometrika 49: 41-56. 1962.

24. Slutsky, Eugen. Über stochastische Asymptoten und Grenzwerte. Metron 5, No. 3: 3-89. 1925.

25. Sobel, Milton and Wald, Abraham. A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. Annals of Mathematical Statistics 20: 502-522. 1949.

26. Tables of the confluent hypergeometric function $F(\frac{n}{2},\frac{1}{2};x)$ and related functions. United States Department of Commerce National Bureau of Standards Applied Mathematics Series 3. 1949.

27. Tables to facilitate sequential t-tests. United States Department of Commerce National Bureau of Standards Applied Mathematics Series 7. 1951.

28. Wald, Abraham. Sequential analysis. New York, N. Y., John Wiley and Sons, Inc. c1947.

29. Widder, David V. Advanced calculus. 2nd ed. Englewood Cliffs, N. J., Prentice-Hall, Inc. c1961.

VII. ACKNOWLEDGMENTS

The author expresses deep gratitude to Professor C. Philip Cox for judicious counsel and direction during the course of this research and for helpful suggestions given after carefully reading the manuscript.

Mrs. Gretchen Snowden of the Numerical Analysis-Programming Group, Iowa State University, is acknowledged for programming and overseeing the Monte Carlo experiments.

The National Institutes of Health (Training Program in Biometry, GM-34-07) are also acknowledged for the financial support of this dissertation.

VIII. APPENDIX

A. Proof of a Theorem, Corollary, and Two Lemmas

Theorem 1:

If

$$\text{(i)} \quad A_{ni} = O(n^{-1/2}), \quad (i=1,2,\dots,r)$$

$$\text{(ii)} \quad E(B_{ni}^2) = O(n^{-1}), \quad (i=1,2,\dots,s)$$

and

$$\text{(iii)} \quad n^{-1}C_n \xrightarrow{(p)} 0 ,$$

then

$$\left(\prod_{i=1}^{r} A_{ni}\right)\left(\prod_{i=1}^{s} B_{ni}\right) C_n = Y_n \xrightarrow{(p)} 0 .$$

Proof:

Given any Δ, $\lambda > 0$, we must show

$$P(\,|Y_n| < \Delta) \geq 1-\lambda \tag{8.1}$$

for n sufficiently large.

Condition (i) is that there exists an $M_i > 0$ such that

$$|A_{ni}| < M_i n^{-1/2} \tag{8.2}$$

for all n.

Condition (ii) is that there exists a $K_i > 0$ such that

$$E(B_{ni}^2) < K_i n^{-1} \tag{8.3}$$

for all n. Let

$$K_i' = K_i, \text{ if } K_i \geq \frac{s+1}{\lambda},$$
$$= K, \text{ otherwise,} \qquad (8.4)$$

where

$$K \geq \frac{s+1}{\lambda} \quad .$$

Then

$$E(B_{ni}^2) < K_i n^{-1} \leq K_i' n^{-1} \qquad (8.5)$$

for all n.

Condition (iii) is that, given $\Delta, \lambda > 0$,

$$P(| n^{-1} C_n | < \Delta) \geq 1-\lambda \qquad (8.6)$$

for n sufficiently large.

Let

$$R_i = \left[\; | A_{ni} | < M_i n^{-1/2} \right], \quad (i=1,2,\ldots,r) \qquad (8.7)$$

where our notation denotes that R_i is the event (set) that (in which) $| A_{ni} | < M_i n^{-1/2}$. Similarly, let

$$S_i = \left[\; | B_{ni} | < K_i' n^{-1/2} \right], \quad (i=1,\ldots,s) \qquad (8.8)$$

and let

$$T = \left[\; | C_n | < \frac{n^{\frac{r+s}{2}} \Delta}{(\prod_{i=1}^{r} M_i)(\prod_{i=1}^{s} K_i')} \right]. \qquad (8.9)$$

Then the event denoting simultaneous occurrence of events $R_1,\ldots,R_r,S_1,\ldots,S_s,T$ is the intersection of these events which we denote as

$$V = \left(\prod_{i=1}^{r} R_i\right)\left(\prod_{i=1}^{s} S_i\right)T \tag{8.10}$$

Letting

$$W = \left[\, |Y_n| < \Delta \,\right] \tag{8.11}$$

we see that

$$V \Longrightarrow W$$

so that

$$P(W) \geq P(V) \ .$$

We wish to show that

$$P(W) \geq 1-\lambda$$

for n sufficiently large.

Letting an asterisk denote the complement (for example, V^* denotes the complement of V), we have

$$P(W) \geq P(V) = 1-P(V^*)$$

$$= 1-P\left(\sum_{i=1}^{r} R_i^* + \sum_{i=1}^{s} S_i^* + T^*\right)$$

$$\geq 1-\left[\sum_{i=1}^{r} P(R_i^*)+\sum_{i=1}^{s} P(S_i^*)+P(T^*)\right] . \tag{8.12}$$

Now

$$P(R_i) = P(\,|A_{ni}| < M_i n^{-1/2}),$$

by 8.7,

$$= 1,$$

by 8.2, so that

$$P(R_i^*) = 1-P(R_i) = 0. \tag{8.13}$$

And,

$$P(S_i) = P(|B_{ni}| < K_i' n^{-1/2}) ,$$

by 8.8,

$$= P(B_{ni}^2 < K_i'^2 n^{-1})$$

$$\geq 1 - \frac{E(B_{ni}^2)}{K_i'^2 n^{-1}} ,$$

by Tchebycheff's inequality,

$$\geq 1- \frac{K_i' n^{-1}}{K_i'^2 n^{-1}}$$

by 8.5,

$$= 1 - \frac{1}{K_i'} ,$$

$$\geq 1 - \frac{\lambda}{s+1} , \tag{8.14}$$

by 8.4. Thus

$$P(S_i^*) = 1 - P(S_i) \leq \frac{\lambda}{s+1} . \tag{8.15}$$

Further,

$$P(T) = P\left[|n^{-1}C_n| < \frac{n^{\frac{r+s-2}{2}} \Delta}{(\prod_{i=1}^{r} M_i)(\prod_{i=1}^{s} K_i')}\right]$$

by 8.9,

$$\geq P\left[\,|n^{-1}C_n| < \frac{\Delta}{\left(\prod_{i=1}^{r} M_i\right)\left(\prod_{i=1}^{s} K_i'\right)}\right] \quad ,$$

$$\geq 1 - \frac{\lambda}{s+1} \tag{8.16}$$

for n sufficiently large, since $n^{-1}C_n \xrightarrow{(p)} 0$.

Therefore

$$P(T^*) = 1 - P(T) \leq \frac{\lambda}{s+1} \quad , \tag{8.17}$$

for n sufficiently large.

Finally — from Equations 8.12, 8.13, 8.15, and 8.17 — we obtain

$$P(W) \geq 1 - \frac{s\lambda}{s+1} - \frac{\lambda}{s+1}$$

$$= 1 - \lambda ,$$

for n sufficiently large, thereby completing the proof.

We now state and prove a corollary to Theorem 1.

Corollary:

Theorem 1 holds for $r=0$, $s\geq 2$ and for $r\geq 2$, $s=0$; i.e. if we allow r or s to be zero in Theorem 1, the theorem remains true for $r+s\geq 2$.

A formal proof of this corollary is not required, because the proof follows easily from that for Theorem 1 by merely considering the two cases $r=0$, $s\geq 2$ and $r\geq 2$, $s=0$ separately.

The two lemmas are now given.

Lemma 1:

If

$$\text{(i)} \quad D_n \xrightarrow{(p)} 0$$

and

$$\text{(ii)} \quad E_n \xrightarrow{(p)} 0 ,$$

then

$$D_n E_n \xrightarrow{(p)} 0 .$$

Lemma 2:

If

$$A_i \xrightarrow{(p)} 0 , \quad (i=1,2,\ldots,k)$$

then

$$\sum_{i=1}^{k} A_i \xrightarrow{(p)} 0 .$$

Proofs for Lemma 1 and Lemma 2 may be developed in a manner similar to that used in proving Theorem 1. However, the lemmas follow as special cases of a more general proposition due to Slutsky (24). Slutsky showed that if $Y_{ni}(i=1, 2,\ldots,m)$ are random variables which converge in probability to the constants $y_i(i=1,2,\ldots,m)$ respectively, then any rational function $R(Y_{n1},Y_{n2},\ldots,Y_{nm})$ converges in probability to the constant $R(y_1,y_2,\ldots,y_m)$, provided that this constant is finite.

B. Maximum Likelihood Estimators

Let

$$L = \log \prod_{i=1}^{n} f(x_i,y_i;\mu,\beta,\sigma,\sigma_x)$$

$$=-n(\log 2\pi+\log \sigma_x+\log \sigma)-\frac{\Sigma x^2}{2\sigma_x^2}-\frac{\Sigma(y-\mu-\beta x)^2}{2\sigma^2} \quad . (8.18)$$

By differentiation, we obtain in the usual way

$$\hat{\mu} = \bar{y} - \hat{\beta}\bar{x} ,$$

$$\hat{\sigma}_x^2 = \frac{\Sigma x^2}{n} ,$$

$$\hat{\beta} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{S_{xy}}{S_{xx}} ,$$

and

$$\hat{\sigma}^2 = \frac{\Sigma(y-\hat{\mu}-\hat{\beta}x)^2}{n} = \frac{1}{n}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) , \tag{8.19}$$

C. Expectation and Variance of $\hat{\beta}, \hat{\sigma}^2$

From 1.12 and 1.13 we have

$$y_j \mid x_j = \mu + \beta x_j + e_j \; ; \quad e_j \sim NI(e;0,\sigma^2) . \tag{8.20}$$

We consider first the expectation of $\hat{\beta}$. By 8.19

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} ;$$

therefore we have

$$\hat{\beta} \mid x = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} \mid x$$

$$= \frac{\Sigma(x-\bar{x})[\beta(x-\bar{x}) + (e-\bar{e})]}{\Sigma(x-\bar{x})^2} \mid x ,$$

by 8.20,

$$= \beta + \beta_e \mid x , \tag{8.21}$$

where

$$\beta_e = \frac{S_{xe}}{S_{xx}} , \tag{8.22}$$

and it is seen that

$$E(\beta_e \mid x) = 0 \ . \tag{8.23}$$

Hence, from 8.21 and 8.23, we obtain

$$E(\hat{\beta}) = E[E(\hat{\beta} \mid x)] = \beta \ . \tag{8.24}$$

We shall now use the following lemma:

Lemma 3:

$$\text{Var}(z) = E_w[V(z \mid w)] + V_w[E(z \mid w)] \ .$$

Here, where z and w denote random variables, the notation $E_w[V(z \mid w)]$ denotes the expectation of $V(z \mid w)$, the expectation being taken over w and $V_w[E(z \mid w)]$ denotes the variance of $E(z \mid w)$, $E(z \mid w)$ being a function of w. The proof of this lemma may be found in the literature, perhaps, but is given in Section D of this chapter for completeness.

By Lemma 3, we have

$$V(\hat{\beta}) = E[V(\hat{\beta} \mid x)] + V[E(\hat{\beta} \mid x)]$$

$$= E[V(\hat{\beta} \mid x)] + 0 \ ,$$

by 8.21 – 8.23,

$$= E\left[E\left[[\hat{\beta}-E(\hat{\beta})]^2 \middle| x\right]\right]$$

$$= E[E(\frac{S_{xe}^2}{S_{xx}^2} \mid x)]$$

by 8.21 and 8.22

$$= \sigma^2 \, E(\frac{1}{S_{xx}}) \ , \tag{8.25}$$

by 8.20.

In order to evaluate $E(\frac{1}{S_{xx}})$ we recall that $x_i \sim NI(x;0,\sigma_x^2)$. Therefore S_{xx}/σ_x^2 is distributed as chi-square with n-1 degrees of freedom. Further, if U,V are independent random variables where uU follows the chi-square distribution with u degrees of freedom and vV is distributed as chi-square with v degrees of freedom, then U/V follows the F-distribution with u,v degrees of freedom ($F_{u,v}$). Then because

$$\frac{v}{v-2} = E(F_{u,v}) = E(\frac{U}{V})$$

$$= E(U)\ E(\frac{1}{V}) \ ,$$

by independence,

$$= \frac{1}{u}\ E(uU)E(\frac{1}{V})$$

$$= E(\frac{1}{V}) \ ,$$

we have

$$E(\frac{1}{V}) = \frac{v}{v-2} \ , \tag{8.26}$$

i.e.

$$E(\frac{1}{vV}) = \frac{1}{v-2} \tag{8.27}$$

Hence it follows that

$$E(\frac{1}{S_{xx}}) = \frac{1}{\sigma_x^2}\ E(\frac{1}{S_{xx}/\sigma_x^2}) = \frac{1}{\sigma_x^2(n-3)} \ , \tag{8.28}$$

and from 8.25 we obtain

$$V(\hat{\beta}) = \frac{\sigma^2}{\sigma_x^2(n-3)} = O(n^{-1}) \tag{8.29}$$

Next we consider the expectation of $\hat{\sigma}^2$. It is shown that the unbiased estimate of σ^2 is

$$\hat{\sigma}^2_{(\text{unbiased})} = \frac{n}{n-2}\,\hat{\sigma}^2 = \frac{1}{n-2}\,(S_{yy} - \frac{S_{xy}^2}{S_{xx}})$$

$$= \frac{1}{n-2}\,(S_{yy} - \hat{\beta}\,S_{xy}) \tag{8.30}$$

We have, upon substituting from 8.20 and 8.21,

$$E[\frac{1}{n-2}(S_{yy}-\hat{\beta}S_{xy})] = \frac{1}{n-2}E\Big[E[(S_{yy}-\hat{\beta}S_{xy}) \mid x]\Big]$$

$$= \frac{1}{n-2}\,E\Big[E[(S_{ee}-\beta_e^2 S_{xx}) \mid x]\Big]$$

$$= \frac{1}{n-2}\,E[(n-1)\sigma^2 - S_{xx}V(\beta_e \mid x)]$$

$$= \frac{1}{n-2}\,E[(n-1)\sigma^2 - S_{xx}\,\frac{\sigma^2}{S_{xx}}]$$

$$= \frac{1}{n-2}\,E[(n-2)\sigma^2] = \sigma^2. \tag{8.31}$$

Next we seek to find $V[\hat{\sigma}^2_{(\text{unbiased})}]$. We have

$$V[\hat{\sigma}^2_{(\text{unbiased})}] = E\left[V[\hat{\sigma}^2_{(\text{unbiased})} \mid x]\right] + V[E(\hat{\sigma}^2_{(\text{unbiased})} \mid x)]$$

$$= E\left[V[\hat{\sigma}^2_{(\text{unbiased})} \mid x]\right] ,$$

since $E(\hat{\sigma}^2_{(\text{unbiased})} \mid x) = \sigma^2$, by 8.31,

$$= E\left[\left(\frac{1}{n-2}\right)^2 V[(S_{yy} - \hat{\beta} S_{xy}) \mid x]\right]$$

$$= \left(\frac{1}{n-2}\right)^2 E\left[V[(S_{ee} - \beta_e^2 S_{xx}) \mid x]\right] ,$$

on substituting from 8.20 and 8.21,

$$= \left(\frac{1}{n-2}\right)^2 E\left[V[(S_{ee} - \frac{S_{xe}^2}{S_{xx}}) \mid x]\right]$$

$$= \left(\frac{1}{n-2}\right)^2 E\left[V(S_{ee} \mid x) + V\left(\frac{S_{xe}^2}{S_{xx}} \mid x\right) - 2 \operatorname{Cov}[(S_{ee}, \frac{S_{xe}^2}{S_{xx}}) \mid x]\right] . \tag{8.32}$$

Now,

$$V(S_{ee} \mid x) = 2\sigma^4 (n-1) , \tag{8.33}$$

[c.f. Kendall and Stuart (17, pp. 60, 277)].

Also,

$$V\left(\frac{S_{xe}^2}{S_{xx}} \mid x\right) = \left(\frac{1}{S_{xx}}\right)^2 V\left[[\Sigma(x-\bar{x})e]^2 \mid x\right]$$

$$= \frac{1}{S_{xx}^2}\left[E([\Sigma(x-\bar{x})e]^4|x) - [E([\Sigma(x-\bar{x})e]^2|x)]^2\right]$$

$$= \frac{1}{S_{xx}^2}[P - Q^2] , \tag{8.34}$$

where

$$P = E([\Sigma(x-\bar{x})e]^4|x) , \tag{8.35}$$

and

$$Q = E([\Sigma(x-\bar{x})e]^2|x) . \tag{8.36}$$

We obtain

$$\begin{aligned}[\Sigma(x-\bar{x})e]^4 &= \Sigma(x-\bar{x})^4e^4 + 4\sum_{i\neq j}(x_i-\bar{x})^3(x_j-\bar{x})e_i^3e_j \\ &+ 3\sum_{i\neq j}(x_i-\bar{x})^2(x_j-\bar{x})^2e_i^2e_j^2 \\ &+ 6\sum_{i\neq j\neq k}(x_i-\bar{x})^2(x_j-\bar{x})(x_k-\bar{x})e_i^2e_je_k \\ &+ \sum_{i\neq j\neq k\neq l}(x_i-\bar{x})(x_j-\bar{x})(x_k-\bar{x})(x_l-\bar{x})e_ie_je_ke_l.\end{aligned} \tag{8.37}$$

Hence, from 8.35 and 8.37 we have

$$\begin{aligned}P &= \Sigma(x-\bar{x})^4(3\sigma^4) + 3\sum_{i\neq j}(x_i-\bar{x})^2(x_j-\bar{x})^2\sigma^4 \\ &= 3\sigma^4[\Sigma(x-\bar{x})^2]^2.\end{aligned} \tag{8.38}$$

Further,

$$[\Sigma(x-\bar{x})e]^2 = \Sigma(x-\bar{x})^2e^2 + \sum_{i\neq j}(x_i-\bar{x})(x_j-\bar{x})e_ie_j \qquad (8.39)$$

so that

$$Q = \Sigma(x-\bar{x})^2\sigma^2\ , \qquad (8.40)$$

and we finally have, from 8.34, 8.38, and 8.40,

$$V(\frac{S_{xe}^2}{S_{xx}}\,|\,x) = \frac{1}{S_{xx}^2}\ (3\sigma^4S_{xx}^2 - \sigma^4S_{xx}^2)$$

$$= 2\sigma^4. \qquad (8.41)$$

In addition it is seen that

$$Cov[(S_{ee},\frac{S_{xe}^2}{S_{xx}})\,|\,x]=E(\frac{S_{ee}S_{xe}^2}{S_{xx}}|x) - E(S_{ee}|x)E(\frac{S_{xe}^2}{S_{xx}}|x). \qquad (8.42)$$

Here,

$$\frac{S_{ee}S_{xe}^2}{S_{xx}} = \frac{1}{S_{xx}}\ \Sigma(e-\bar{e})^2S_{xe}^2$$

$$= \frac{1}{S_{xx}}\ [\Sigma e^2 - \frac{(\Sigma e)^2}{n}]S_{xe}^2$$

$$= \frac{1}{S_{xx}}[\Sigma e^2S_{xe}^2 - \frac{(\Sigma e)^2}{n}S_{xe}^2] \qquad (8.43)$$

Also,

$$\Sigma e^2S_{xe}^2 = \Sigma e^2[\Sigma(x-\bar{x})^2e^2 + \sum_{i\neq j}(x_i-\bar{x})(x_j-\bar{x})e_ie_j]$$

$$= \Sigma(x-\bar{x})^2e^4 + \sum_{i\neq j}\ e_i^2(x_j-\bar{x})^2e_j^2$$

$$+2\sum_{i\neq j}(x_i-\bar{x})(x_j-\bar{x})e_i^3e_j + \sum_{i\neq j\neq k}(x_j-\bar{x})(x_k-\bar{x})e_i^2e_je_k, \qquad (8.44)$$

and

$$(\Sigma e)^2 S_{xe}^2 = \Sigma(x-\bar{x})^2 e^4 + \sum_{i\neq j} e_i^2 (x_j-\bar{x})^2 e_j^2$$

$$+2\sum_{i\neq j}(x_i-\bar{x})(x_j-\bar{x})e_i^3 e_j + \sum_{i\neq j\neq k}(x_j-\bar{x})(x_k-\bar{x})e_i^2 e_j e_k$$

$$+2\sum_{i\neq j}(x_i-\bar{x})^2 e_i^3 e_j + \sum_{i\neq j\neq k}(x_i-\bar{x})^2(x_j-\bar{x})(x_k-\bar{x})e_i^2 e_j e_k$$

$$+2\sum_{i\neq j}(x_i-\bar{x})(x_j-\bar{x})e_i^2 e_j^2 + 4\sum_{i\neq j\neq k}(x_i-\bar{x})(x_k-\bar{x})e_i^2 e_j e_k$$

$$+\sum_{i\neq j\neq k\neq l}(x_k-\bar{x})(x_l-\bar{x})e_i e_j e_k e_l. \qquad (8.45)$$

Then, Equations 8.43 — 8.45 give

$$E\left(\frac{S_{ee}S_{xe}^2}{S_{xx}}\middle|x\right) = \frac{1}{S_{xx}}\Big[S_{xx}3\sigma^4+(n-1)\sigma^4 S_{xx}-\frac{1}{n}[S_{xx}3\sigma^4$$

$$+(n-1)\sigma^4 S_{xx}+2\sum_{i\neq j}(x_i-\bar{x})(x_j-\bar{x})\sigma^4]\Big]$$

$$= \frac{1}{S_{xx}}\Big[3\sigma^4 S_{xx}+(n-1)\sigma^4 S_{xx}-\frac{1}{n}[3\sigma^4 S_{xx}$$

$$+(n-1)\sigma^4 S_{xx} - 2\sigma^4 S_{xx}]\Big]$$

$$= (n+1)\sigma^4 . \qquad (8.46)$$

Also,

$$E(S_{ee} \mid x) = (n-1)\sigma^2 \tag{8.47}$$

and

$$E(\frac{S_{xe}^2}{S_{ee}} \mid x) = E(S_{xx}\beta_e^2 \mid x)$$

$$= S_{xx}V(\beta_e \mid x) = \sigma^2 . \tag{8.48}$$

So, from 8.42, 8.46, and 8.48 we obtain

$$Cov[(S_{ee},\frac{S_{xe}^2}{S_{xx}}) \mid x] = (n+1)\sigma^4-(n-1)\sigma^4$$

$$= 2\sigma^4 . \tag{8.49}$$

Hence, Equations 8.32, 8.33, 8.41, and 8.49 give

$$V[\hat{\sigma}^2_{(unbiased)}] = \frac{[2\sigma^4(n-1)+2\sigma^4-4\sigma^4]}{(n-2)^2}$$

$$= \frac{2\sigma^4}{n-2} = O(n^{-1}) . \tag{8.50}$$

D. Proof of Lemma 3

Lemma 3:

$$V(z) = E_w[V(z \mid w)] + V_w[E(z \mid w)] .$$

Proof:

$$V(z) = E[z-E(z)]^2$$

$$= E[z-E(z \mid w)]^2 + E[E(z \mid w) - E(z)]^2$$

$$+ 2E[z-E(z \mid w)][E(z \mid w)-E(z)] .$$

Now

$$E[z-E(z \mid w)]^2 = E_w E\Big[[z-E(z \mid w)]^2 \Big| w\Big] = E_w[V(z \mid w)],$$

$$E[E(z \mid w)-E(z)]^2 = E_w E\Big[[E(z \mid w)-E_w E(z \mid w)]^2 \Big| w\Big]$$

$$= E_w[E(z \mid w)-E_w E(z \mid w)]^2$$

$$= V_w[E(z \mid w)] \ ,$$

and

$$E[z-E(z \mid w)][E(z \mid w)-E(z)]$$

$$= E_w E\Big[[z-E(z \mid w)][E(z \mid w)-E(z)] \Big| w\Big]$$

$$= E_w[E(z \mid w)-E(z \mid w)][E(z \mid w)-E(z)] = 0 \ ,$$

giving the lemma.